SOLID-PHASE SYNTHESIS OF C-TERMINAL PEPTIDE LIBRARIES FOR STUDYING THE SPECIFICITY OF PROTEIN FARNESYLTRANSFERASE

A THESIS SUBMITTED TO THE FACULTY OF THE UNIVERSITY OF MINNESOTA BY

Yen-Chih Wang

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

Advisor: Dr. Mark D. Distefano

July, 2014

© Yen-Chih Wang 2014

Table of Contents

List of Abbreviations

BCIP, 5-Bromo-4-chloro-3-indolyl phosphate disodium salt

BOP, (benzotriazol-1-yloxy)tris-(dimethylamino)phosphonium hexafluorophosphate

BPB, bromophenol blue

CCA, α-cyano-4-hydroxycinnamic acid

CDI, 1,1′-carbonyldiimidazole

CDT, 1,1′-carbonyl-di-(1,2,4-triazole)

6-Cl-HOBt, 6-chloro-1-hydroxy-1H-benzotriazole

DCC, dicyclohexylcarbodiimide;

Dde, 1-(4,4-dimethyl-2,6-dioxacyclohexylidene)ethyl

DIC, *N,N'*-diisopropylcarbodiimide

DIEA, N,N-diisopropylethylamine

Dmab,

2-{1-[4-(Hydroxymethyl)phenylamino]-3-methylbutylidene}-5,5-dimethyl-1,3-cycloh exanedione

DMAP, dimethylamino-pyridine

DTT, dithiothreitol

FPP, farnesyl pyrophosphate

HCTU, O-(6-Chloro-1-hydrocibenzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate

HMPA, 4-hydroxymethyl-phenoxyacetic acid

PFTase, protein farnesyltransferase

PPTS, pyridinium p-toluenesulfonate

SP-AP, streptavidin alkaline phosphatase

TBAF, tetrabutylammonium fluoride

TBTA, tris[(1-benzyl-1*H*-1,2,3-triazol-4-yl)methyl]amine

TBDMS. tert-butyldimethylsilyl ether

TCEP, tris(2-carboxyethyl)phosphine hydrochloride

TFA, trifluoroacetic acid

Tris, tris(hydroxymethyl)aminomethane

List of Tables

List of Figures

Chapter 1: A Review of Synthesis and Screening Peptide Libraries with Free C-termini

1.1 Introduction to peptide library synthesis

(a) Solid-phase peptide synthesis

Combinatorial synthesis has been one of the most rapidly developing fields in the academic laboratory and pharmaceutical industry in the past two decades and is now seen as an essential tool in both the discovery and the development of new bioactive molecules and drugs. Put at its simplest, combinatorial synthesis is a means of producing a large number of compounds in a short period of time, using a defined reaction route and a large variety of starting materials and reagents. Usually, this is done on a small scale using solid phase synthesis, so that the process can be automated or semi-automated. This allows each reaction of the synthetic route to be carried out in several reaction vessels (or spots) at the same time and under identical conditions, but using different reagents for each vessel (or each spot).

Solid phase synthesis was pioneered for the synthesis of peptides, and most of the early work carried out on combinatorial synthesis was performed on peptides as well because the building blocks (amino acids) have some regular functionalities (amino group and carboxylic acid group). There are already 20 amino acids available

from Nature meaning that there are already many building blocks to generate mixtures, which will be referred to combinatorial libraries in this manuscript. For example, if a library of tetrapeptides is to be synthesized with 20 variations for each amino acid, a mixture of $20^4 = 1.6\times10^5$ kinds of peptides will be obtained at the end. If unnatural amino acids are incorporated for library build-up, a library of bigger size will be generated. Peptide libraries will be the focus of this manuscript.

The idea of chemical synthesis of peptides is that the first amino acid is attached to a solid support through its carboxyl group and then each N-protected amino acid is added in turn. During the coupling, the carboxyl group of the new coming amino acid must be activated. The commonly used activating reagents are a carbodiimide (for example, DIC), a uranium (for example, HCTU) or a phosphonium (for example, BOP) (Figure 1.1). After each addition, the N-protection must be removed before the next amino acid is added. The commonly used N-protecting group is Fmoc which can be removed under basic condition such as 20% piperidine so this process is also called Fmoc-solid phase peptide synthesis (Fmoc-SPPS). The growing peptide chain is attached to the solid support so that all waste products, removed protecting groups, excess reagents and inorganic rubbish can be washed out after each operation.

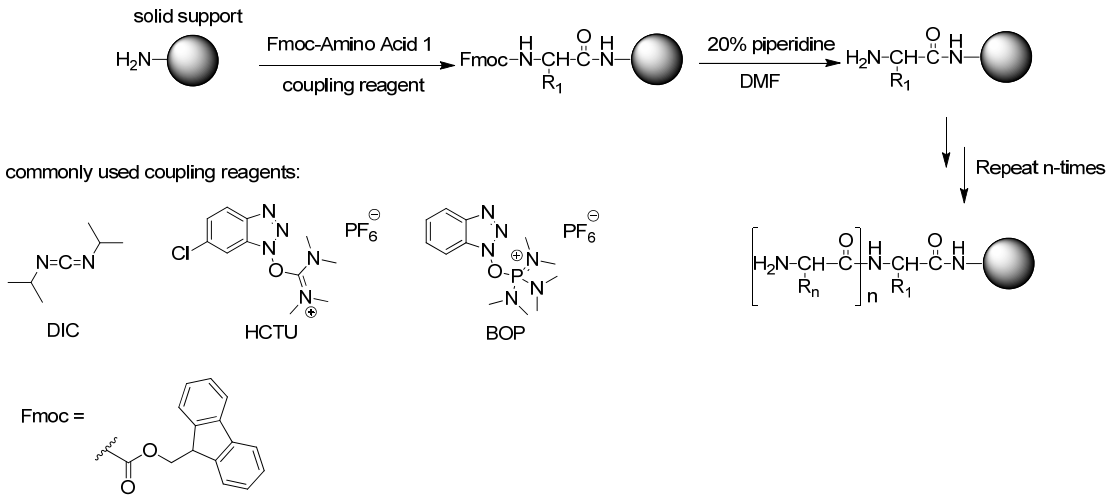


Figure 1.1 A schematic view of Fmoc-solid phase peptide synthesis.

Combinatorial synthesis can be carried out such that a single product is obtained in each different reaction flask—a process known as parallel synthesis. Alternatively, the process can be designed such that mixtures of compounds are produced in each reaction vessel.

Parallel synthesis can be performed on resins which is a cross-linked insoluble polymeric bead or on pieces of chemically functionalized papers which will be referred to membranes in this manuscript. One manual approach to parallel synthesis is called teabag procedure.[1] The polymeric support resin is sealed in labeled meshed containers (known as teabags). The teabags are then placed in bottles which act as the reaction vessels (Figure 1.2). In the case of a peptide synthesis, the first amino acid is added to the resin (different amino acids for different bottles). All teabags in one specific bottle now have the same amino acid linked to the resin. The teabags from

every bottle are now combined in one vessel for deprotection and washing. The teabags can then be redistributed between the bottles for the addition of a second amino acid, recombined for deprotection and washing, redistributed for addition of the next amino acid, and so on.

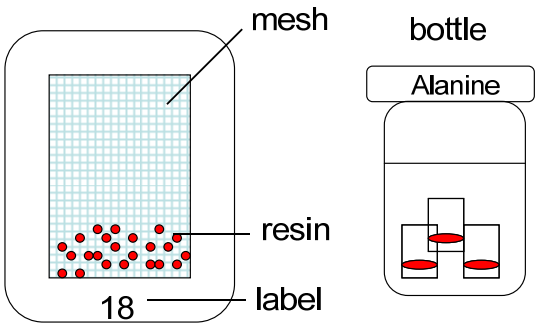


Figure 1.2 Teabag procedure.

Another approach to parallel synthesis is called SPOT synthesis. In SPOT synthesis, the solutions of activated amino acids are positionally addressed and delivered in small drops to distinct points on the solid surface such as functionalized cellulose membrane or glass forming a pattern of small spots. Peptides arrays are synthesized in a stepwise manner following the standard Fmoc-based peptide chemistry (Figure 1.3).[2] Using the automated SPOT synthesis, it is routine to synthesize and screen 600 peptides on the membrane of 10×15 cm^2 size.

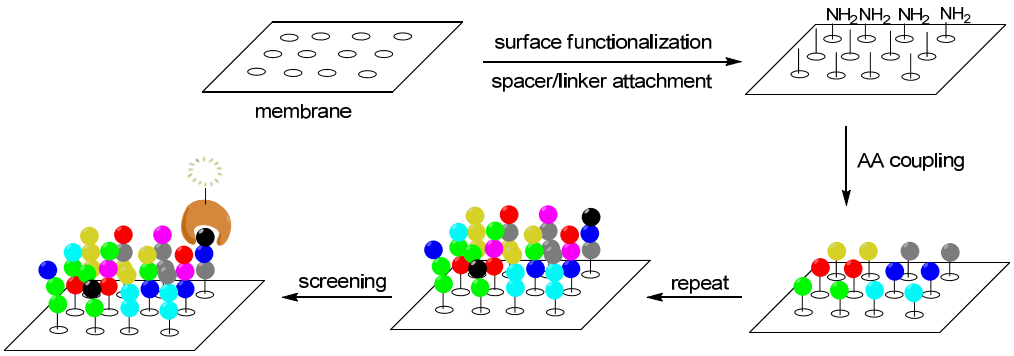


Figure 1.3 A schematic view of the SPOT synthesis and their applications in the binding assays.

The disadvantage of the parallel synthesis is that the resulting library has a small size — within 1000 members is usually made each time. If a bigger library is to be synthesized, the split and mix method should be a better choice. The solid phase of split and mix method is a resin (bead). The split and mix method consists of three steps[3]: (1) Each amino acid is coupled to resin beads. (2) The beads are mixed together and separated into equal portions. The number of portions depends on how many residues are going to be coupled. Then each portion is reacted with a different amino acid. (3) Repeat the process of (1) and (2). All beads are mixed together and split into portions. Each portion is reacted with one of the building amino acids (Figure 1.4). In Figure 1.4, we have three amino acids (Gly, Ala and Val) as our building blocks and we want to make a tripeptide library. After the split and mix synthesis, we will get all 27 possible tripeptides. Synthesis and screening of longer peptides is routine nowadays. If we have 20 amino acids and we want to make a

pentapeptide library, a library size of $20^5 = 3.2\times10^6$ members will be generated. The size is much bigger than what could be made by parallel synthesis. However, in this case you have a mixture of compounds in a vessel. Note that in the split and mix method, each bead has a unique peptide sequence attached to it so the resulting library is also called one-bead-one-compound (OBOC) library. After screening the bead library, the peptide sequences from the selected beads can be determined by Edman degradation,[4] partial Edman degradation—mass spectrometry (PED/MS)[5] or tandem mass spectrometry (MS/MS).[6]

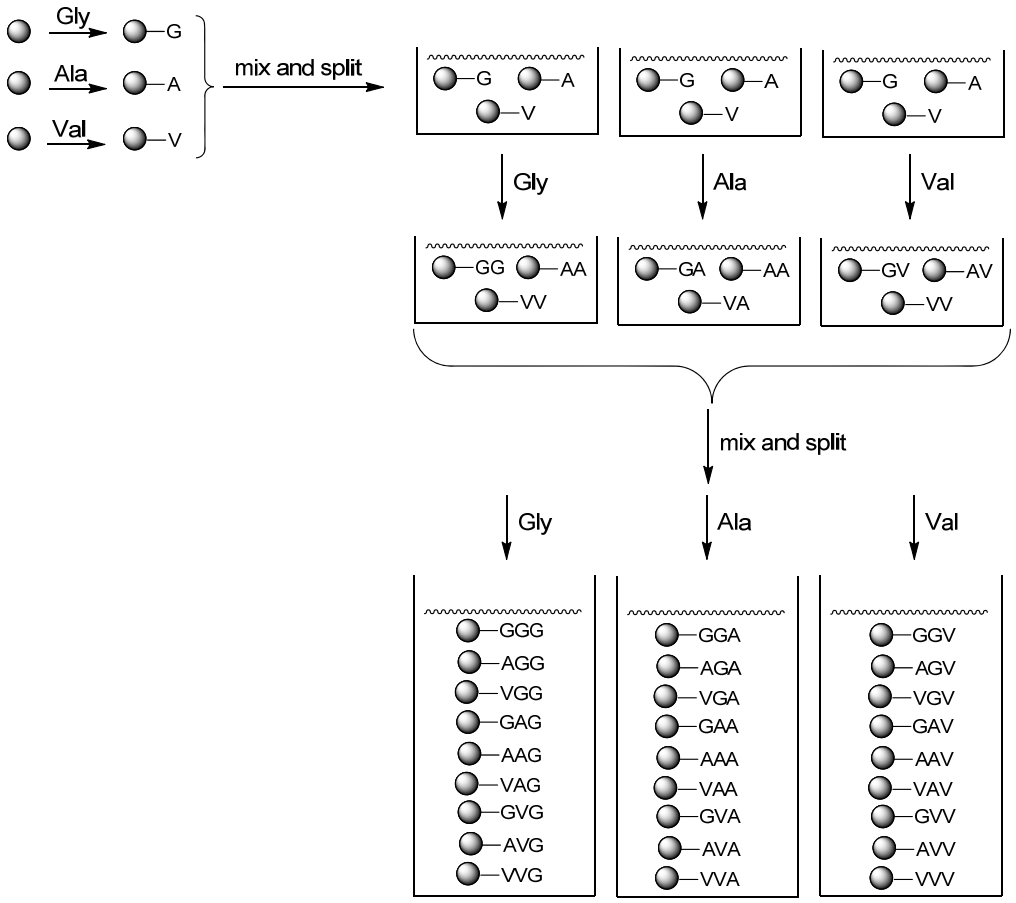


Figure 1.4 A schematic view of the split and mix method (or OBOC method).

(b) Phage display

In contrast to solid-phase peptide synthesis, in biological systems a peptide or protein is synthesized from its N-terminal end to its C-terminal end. In cells the genetic message is transcribed from DNA into RNA and then translated into protein. A sequence of three nucleotide bases, called a codon, specifies a particular amino acid that is to be incorporated into a protein. Each amino acid is specified by a three-base sequence and this relationship is known as the genetic code. For example, UCA on mRNA codes for the amino acid serine, while CAG codes for glutamine. Since in nature the protein sequence is determined by DNA sequence, the problem of how to make peptide libraries is actually the problem of how to make DNA libraries. There are several ways of making DNA libraries. First, degenerate DNA primers can be synthesized by solid-phase synthesis. For example, if a random nonapeptide is to be attached to the C-terminus of a protein, a primer sequence containing $(NNK)_9$ can be made and incorporated at the C-terminus of the protein (Figure 1.5). N stands for any of the 4 nucleotides (A, T, C, G) and K stands for G or T. The reason why NNK sequences for random codons are used instead of NNN is because we want the chance of random codons happening to be stop codons minimized. The codons for the 20 amino acids are still retained in the NNK codons.

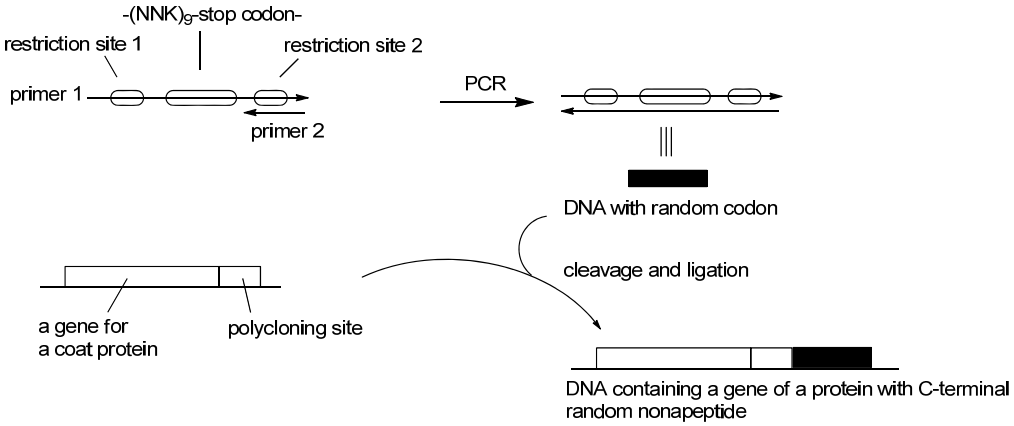


Figure 1.5 Illustration of making a random DNA library by the use of degenerate primer (primer 1).

If we want a library of certain sequences, for example DNA sequences present in human genome, a completely random DNA library may not be the optimal choice. Customized DNA microarray can be ordered. In this case, Oligonucleotides encoding sequences of interest are generated and linked on a microarray slide. The sequences can be amplified by PCR and then cloned into a vector, for example, a phagemid (Figure 1.6).

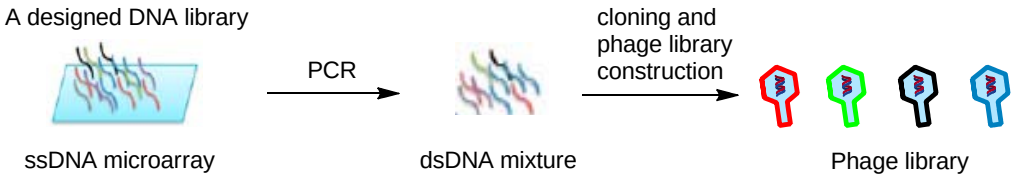


Figure 1.6 A procedure of making a customized DNA library.

A piece of DNA sequence can be randomized by error-prone PCR. In this method, during PCR not only dNTP are added to the solution, but also mutagenic

nucleotides like 8-oxo-2'-deoxyguanosine (8-oxo-dGTP) and 4-dihydro-8H-pyrimido[4,5-c][1,2]oxazine-7-one (dPTP) are added. During DNA replication, the mutagenic nucleotides can mispair with the base of the opposite strand to cause mutations. The rate of mutagenesis can be controlled by the number of PCR cycles and the concentration of mutagenic nucleotides.

Phage display is a method of identifying peptides, which can be recognized by a target such as a receptor or an enzyme, from a library. Phage display depends on the fact that the phages are made up of DNA coated with proteins encoded by that DNA. If a particular peptide sequence is translationally fused to the coding sequence for one of the coat proteins of the phage, all the progeny of that phage will display the particular peptide sequence on their surface. To use this technology, a randomized peptide-coding sequence is fused to a coat protein so that different phages display different versions of the peptide on their surface. The phage that displays a version of the peptide which binds to the target are "panned for" and isolated. If only a few phage particles from the mixture can be separated from most of other phages that express different peptides, the phages can be propagated to make hundreds of millions of identical copies. The cloned DNA can then be sequenced to determine the particular peptide sequence which binds to the target (Figure 1.7). Many different types of phages have been adapted for phage display. In some cases, different coat

proteins from a phage can be chosen to display the peptides. Which phage and which coat protein to use depend on the specific application.

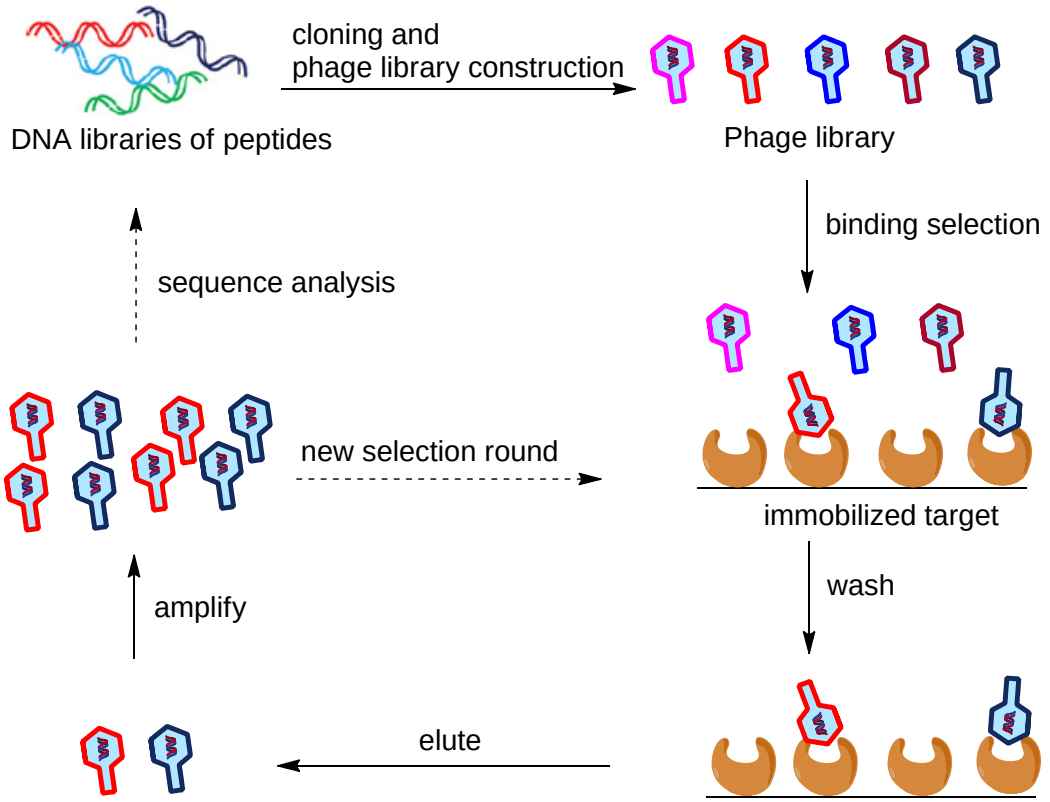


Figure 1.7 Schematic procedure of phage display.

1.2 Synthesis and screening of C-terminal peptide library

(a) Interactions involving peptides with C-termini

Protien-protein interactions and protein post-translational modifications participate in a wide range of cellular process such as the control of protein localization, protein turn-over, enzyme activity and regulation of signal transduction. Misregulation of protein-protein interactions and post-translational modifications

often involve in many diseases such as cancers highlighting their importance in cell physiology. Among these interactions, some proteins bind to their partners at the C-terminus. The PDZ domain (acronym of the synapse-associated protein PSD-95/SAP90, the septate junction protein Discs-large and the tight junction protein ZO-1) is one of the most common protein modules in eukaryotic cells. These domains bind C-terminal motifs of target proteins. It is documented that the last four amino acids in the C-terminal targets are of crucial importance in determining PDZ domain affinity. The residue selectivity in these positions is commonly used to distinguish peptide ligands into classes.[7,8] For example, S/T-X-Ψ is for class I, Φ-X-Ψ for class II, and E/D-X-Ψ for class III with Ψ and Φ represent aliphatic and hydrophobic amino acids, respectively. Other classes of peptides recognized by different PDZ domains were also reported.[9-11] Some proteins harbor multiple PDZ domains that are often arranged in closely linked groups. This renders the proteins capable of binding to several different transmembrane or intracellular partners at the same time, thus organizing multimeric complexes.

Neurophilin-1 (NRP-1) is another example of proteins recognizing the C-termini of their partners. NRP-1 is a cell-surface receptor that plays an important role in angiogenesis and cardiovascular development.[12,13] NPR-1 mediates vascular permeability induced by VEGF and certain semaphorins.[14,15] NPR-1 interacts with

VEGF and semaphoring 3A through their C-terminal basic residues. Several peptides homologous to the VEGF C-terminus are known to compete with VEGF-NPR-1 interaction.[16,17]

After a protein is synthesized, the structure can be modified by other enzymes. For example, a protein can be cleaved by a protease, phosphorylated by a kinase or lipidated by a prenyltransferase. This process is called post-translational modification. The purpose of post-translational modification is to switch on (or switch off) an enzyme's activity, to locate the protein to certain place in the cell or to change the binding affinity to other proteins. Some enzymes carrying out such reactions recognize C-termini of their substrates. For example, prenylation of a variety of proteins with isoprenoid is essential for normal cellular functions and has important roles in numerous diseases. Protein prenylation involves the addition of farnesyl (C15) or geranylgeranyl (C20) isoprenoids near the C-termini of proteins and is catalyzed by protein farnesyltransferase (PFTase) and geranylgeranyltransferases (GGTase), respectively.[18] In a previous model, the enzymes recognize a C-terminal tetrapeptide sequence called Ca_1a_2X motif where C is the cysteine to which the isoprenoid is attached, a_1 and a_2 are usually aliphatic amino acids and X is the major determinant for modification by either PFTase or GGTase I (Figure 1.8).[19,20]

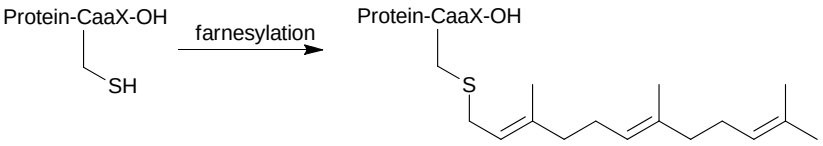


Figure 1.8 C-terminal farnesylation.

Connecting proteins to ubiquitins (Ubs) is another modification associated with a variety of cellular functions, including regulation of protein degradation, control of protein subcellular localization, induction of protein-protein interactions and regulation of enzymatic activities. Ubiquitin is a 76-residue protein that can be linked to other cellular proteins through an isopeptide bond between the C-terminal carboxylate of Ub and the ε-amino group of a Lys residue on the modified protein.[21] The Ub modification pathway involves three enzymes, E1, E2, E3. The E1 enzyme catalyzes the attachment of Ub molecule to a Cys residue of E1 itself to form a Ub-E1 thioester bond. Next, Ub is transferred to an E2 to form a Ub-E2 conjugate via a trans-thioester reaction. Finally an E3 enzyme catalyzed the transfer of Ub from E2 to the substrate protein (Figure 1.9).[22] So far, there are 2 E1s, 50 E2s and more than 1000 E3s identified in the human genome.[23,24] These enzymes assemble overlapping and intersecting networks for Ub transfer to the cellular target proteins.

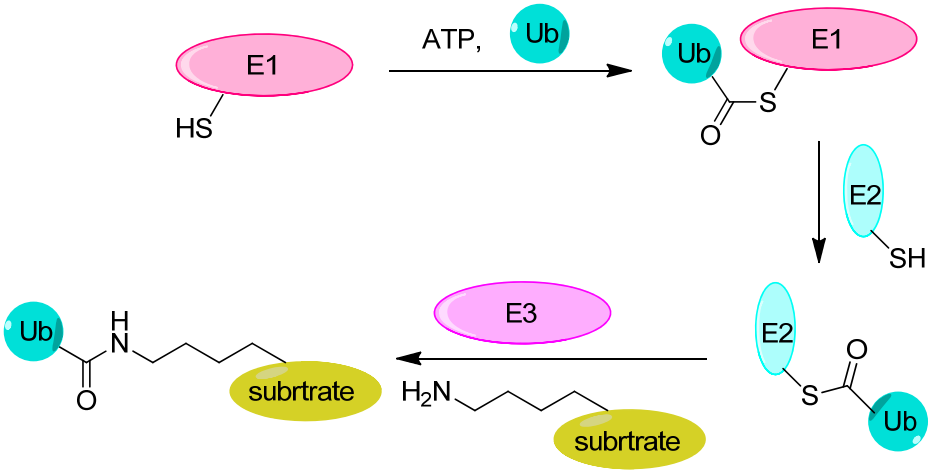


Figure 1.9 The enzymatic cascade of E1-E2-E3 that catalyzes Ub transfer onto substrate proteins.

There are nearly 20 structurally Ub-related proteins in eukaryotes such as Nedd8 and SUMO. These proteins are called ubiquitin-like proteins (UBLs). Although structurally related, UBLs regulate a strikingly diverse set of cellular processes, including nuclear transport, proteolysis, translation, autophagy, and antiviral pathways.[25] The UBL modification involves similar E1-E2-E3 cascade. Several proteins involved in these ubiquitin-related processes need to specifically recognize Ub or UBL to perform their function, such as deubiquitylation enzymes (DUBs), certain trafficking proteins and ubiquitin conjugating enzymes. To recognize Ub or UBL, many proteins utilize a variety of ubiquitin-binding domains.[26] Some domains, like BUZ domain, have been found to interact with ubiquitin by binding to its free C-terminus.[27,28]

Not only there are examples of proteins recognizing C-terminal peptides, but also

some peptide analogues can bind to their targets at their C-termini. Vancomycin is a narrow-spectrum antibiotic produced by a bacterium *Streptomyces orientalis*. Vancomycin is derived biologically from a linear heptapeptide containing 5 aromatic residues. Oxidative coupling, chlorination, hydroxylation and final addition of two sugars complete the structure (Figure 1.10). There is a pocket in the vancomycin structure into which the tail of the bacterial cell wall building block's pentapeptide moiety can fit. It is held there by the formation of five hydrogen bonds between the cell wall building block and vancomycin (Figure 1.10). Because vancomycin is a large molecule, it covers the tails and acts as a steric shield, blocking access to the transglycosidase and transpeptidase enzymes for cell wall construction.[29]

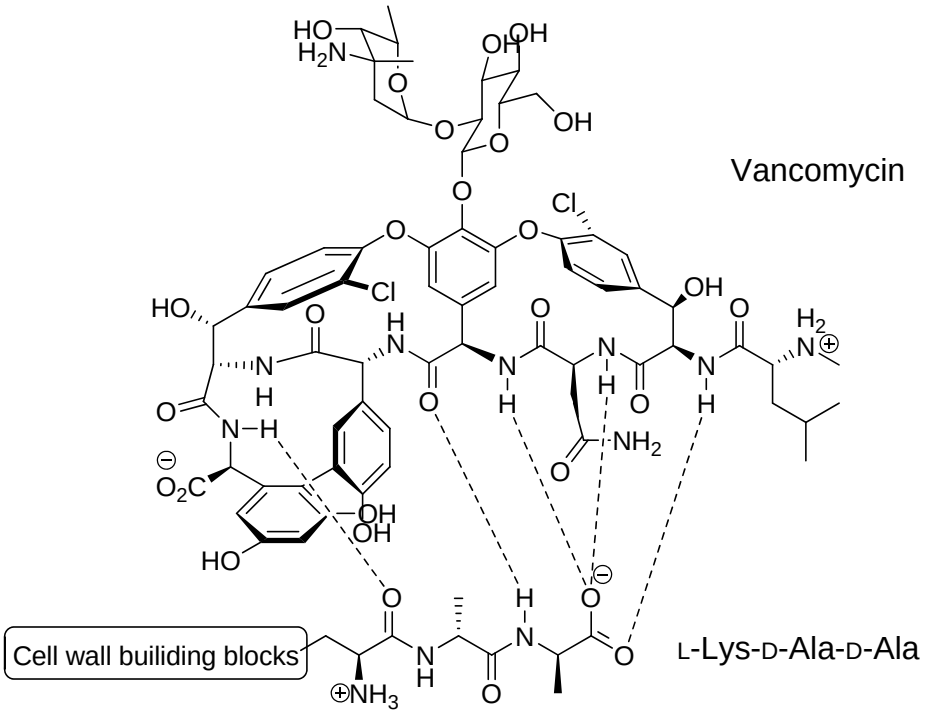


Figure 1.10 Vancomycin and binding interactions to the L-Lys-D-Ala-D-Ala moiety.

(b) Solid-phase peptide synthesis

The well-established solid-phase peptide synthesis starts from C-terminus to N-terminus. The C-terminus of the peptide is tethered to the solid support (Figure 1.1). This technical problem makes the determination of the sequence specificity of targets which require free C-termini for recognition difficult. One way to solve this problem is to make peptides on solid phase but then release the peptides and performed the screening assays in solution. Spaller and coworkers made two chemical libraries of hexapeptide YKQTXV template in a parallel synthesis platform with 92 variations of X in the library I and 186 variations of X in the library II.[30] Resins were arrayed into 96-well filter plate. After the peptides were cleaved from the beads, an ELISA-type assay was used to screen the binding of peptides to third PDZ domain (PDZ3) of PSD-95 protein (Figure 1.11).

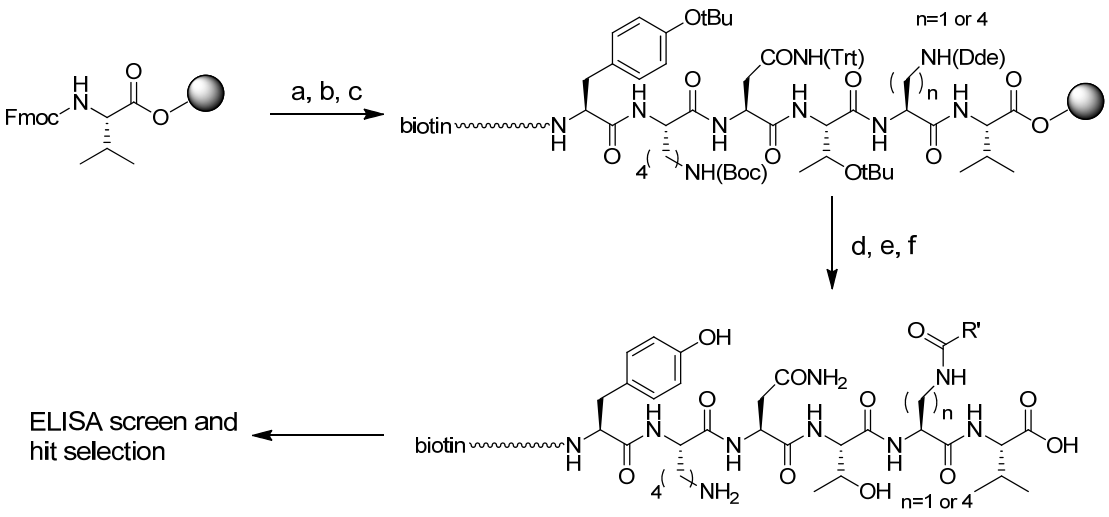


Figure 1.11 Synthesis and screening of C-terminal-displayed organic acid-modified

libraries for PSD-95 PDZ3. Reagents and conditions: (a) Solid-phase peptide synthesis; (b) linker coupling; (c) biotinylation; (d) Dde removal (2% hydrazine); (e) array resin into 96-well filter plates and couple to a variety of organic acids (R'CO_2H); (f) TFA treatment.

Cantley and coworkers used a similar approach to investigate the peptide-binding specificities of nine PDZ domains from various proteins.[7] A primary library, KNXXXXXXXX-COOH, and a secondary library, KNXXXXXX(S,T,Y)XX-COOH, were made by split and mix method. The peptides were released into solutions as a mixture. Some peptides with affinity to a PDZ domain were enriched and then sequenced by Edman degradation. Since a peptide pool was sequenced at the same time, preference for certain amino acids at a given position was obtained, not the individual binding sequences. The above approaches used off-bead screening platform meaning that the peptides are not attached to the solid phase when they are tested for biological activity. However, there are two disadvantages for off-bead screening. First, the off-bead screening is usually slower than on-bead screening. In the on-bead screening platform, 10^8 beads can be readily screened which is faster than the 96-well plate format in the first example. Second, there are cases where the peptides released prove to be insoluble in the test assay and give a false-negative result which could be avoided if the peptides were attached to the bead during the assay.

If on bead screening is to be performed, some ingenuity is going to be taken to present a free carboxyl terminus for target recognition because in a standard

solid-phase peptide library, the carboxyl terminus of the peptide is tethered to the solid support. Many researchers have prepared peptide libraries with free C-termini by a method called “peptide inversion” (Figure 1.12). In this approach, a peptide is first synthesized in the conventional C to N manner. However, during this synthesis, a special handle is installed. This special moiety contains two functional groups (FG 1 and FG 2). Peptide synthesis continues on FG 1 in the traditional C to N manner. Next, the complete peptide (AA_n—AA_1) is cyclized between its N-terminus and the internal FG 2 to give a cyclic peptide. Note that the N-terminus of the complete peptide can be an exposed amino group or another functional group installed (FG 3 in Figure 1.12). Finally, cleavage of the cyclic peptide from C-terminus of AA_1 releases the free C-terminus for screening.

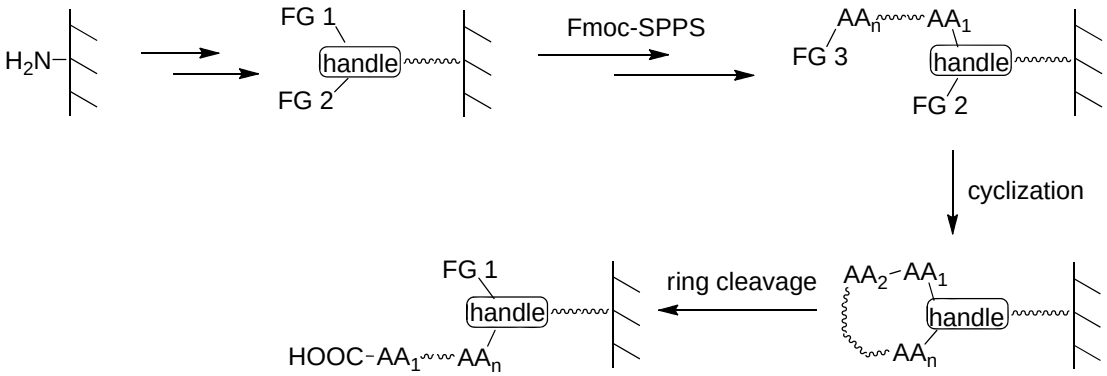


Figure 1.12 Illustration of peptide inversion strategy, a method of Synthesis of C-terminal libraries.

Marlowe can coworkers used this strategy to make resin-bound peptides with

free C-termini (Figure 1.13).[31] Note that in the structure **1.1** during their synthesis, the HMPA-Lys-Suc is the special handle in Figure 1.12. The HMPA is the FG 1 and Suc is the FG 2 which is protected by o-nitrobenzyl group. Peptide synthesis continues from HMPA and the bond between the first amino acid to the HMPA is an ester bond. After the peptide is complete, o-nitrobenzyl group is removed by UV light to expose the free carboxylic acid (FG 2). Intramolecular amide bond formation generates a cyclic peptide. The ether bond between the peptide and HMPA is labile in acidic condition because of the electron-donating oxygen atom directly attached to the phenyl ring. Under 95% TFA, the ether bond between peptide and HMPA is cleaved to give a peptide with C-terminus exposed and N-terminus attached to the surface. Later, those peptides can be processed by carboxypeptidase Y to release the C-terminal 4 amino acids into solutions.

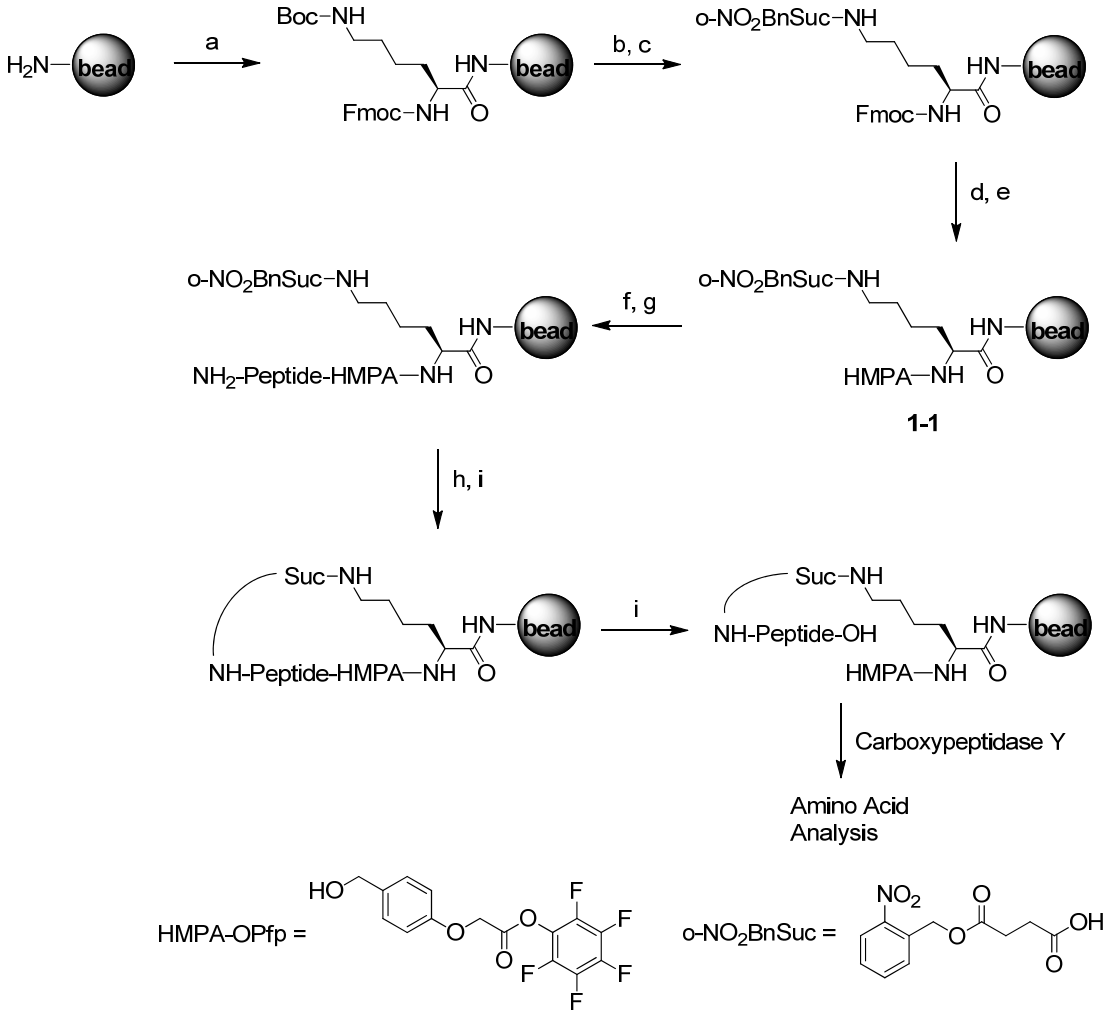


Figure 1.13 Synthesis of resin-bound peptides with free C-termini and enzymatic analysis. Reagents and conditions: (a)Fmoc-Lys(Boc)-OH, DIC, HOBt, CH_2Cl_2; (b) 50% TFA / CH_2Cl_2; (c) o-NO_2BnSuc, DIC, HOBt, DMF; (d) 20% piperidine / DMF; (e) HMPA-OPfp, DIEA, DMF; (f) Fmoc-Amino acid, DIC, HOBt, DMAP, DMF; (g) Solid-phase peptide synthesis; (h) hv at 360 nm, 5% NH_4OH; (i) BOP, DIEA, DMF; (j) 5% anisole, 5% ethanedithiol, 90% TFA.

If peptide libraries are made by the above example, a problem occurs. How do we determine the peptide sequences from the mixtures of beads? Edman degradation is a common way to characterize peptide sequences. However, Edman degradation requires peptides with free N-termini. This problem is avoided if parallel synthesis is

adopted. For example, in SPOT synthesis, the peptide sequences are known based on the locations of spots on a membrane. Volkmer and coworkers used SPOT synthesis to analyze PDZ domain regulation mechanisms (Figure 1.14).[32] In their synthetic method, the Cys-βAla-HMPA in structure **1-2** is the special handle shown in Figure 1.12. The HMPA is the FG 1 and Cys is the FG 2 which is protected by trityl group. Peptide synthesis continues from HMPA and the bond between the first amino acid to the HMPA is an ester bond. The traditional esterification methods are not compatible with SPOT synthesis format due to the long reaction time and volatility of the solvent (CH_2Cl_2 is usually the solvent). They previously developed a new ester bond formation method utilizing CDI (1,1'-carbonyldiimidazole) as the activating reagent for amino acid coupling.[33] The reaction time and conditions have been optimized for SPOT synthesis. CDI was replaced by CDT (1,1'-carbonyl-di-(1,2,4-triazole)) as an activator when the amino acids are Gln, Tyr or Pro to avoid a precipitation problem (step b in Figure 1.14). After the peptide is complete, the N-terminus is reacted with bromoacetic acid. The resulting bromoacetamide is a reactive electrophile (FG 3 in Figure 1.12) which can react with the deprotected Cys under mild basic condition to form a cyclic peptide. Then the peptides with free C-termini can be generated by TFA treatment because of the lability of the ether bond between the HMPA and the amino acid 1. The use of Cys as the nucleophilic FG 2 meaning that Cys cannot be

incorporated in the peptide library synthesis but the use of orthogonally protected Cys[34,35] should solve this problem.

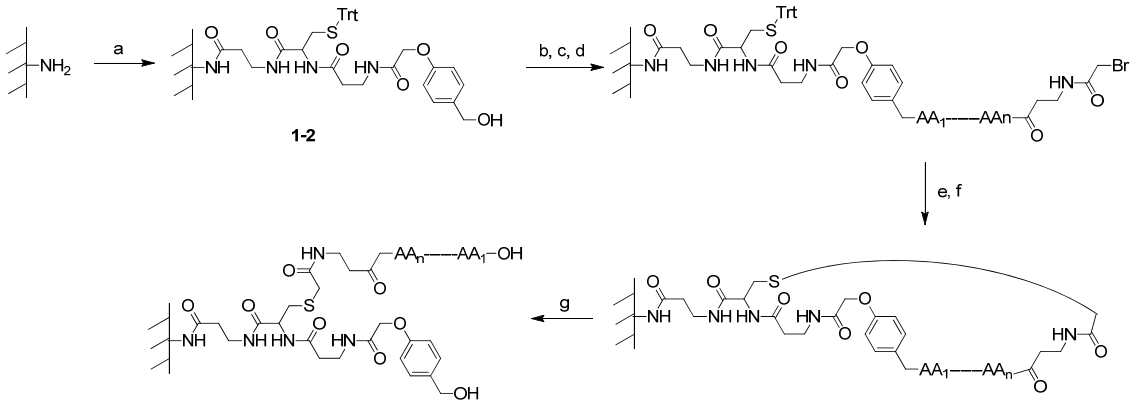


Figure 1.14 Synthesis of peptides with free C-termini by SPOT synthesis for PDZ binding assay. Reagents and conditions: (a) Standard Fmoc-SPPS, β-Ala, Cys(Trt), β-Ala, HMPA; (b) Fmoc-Amino acid, CDI, DMF (the CDI is replaced by CDT when the amino acids are Gln, Tyr or Pro); (c) Standard Fmoc-SPPS; (d) bromoacetic acid 2,4-dinitrophenyl ester, NMP; (e) 7% TFA, CH_2Cl_2; (f) Cs_2CO_3 / H_2O / DMF; (g) 60% TFA followed by 90% TFA.

Volkmer and coworkers determined the specificity of PDZ domains by incubating the membrane in a solution containing PDZ domains with GST fusion. The high affinity peptides were detected by the use of anti-GST antibody with HRP fusion.[32,36,37] Later they used a similar strategy to find candidate substrates of a periplasmic protease containing PDZ domains (Tsp) from *E. coli*.[38] 648 proteins were found to be potential substrates based on bioinformatic analysis. Their C-terminal sequences were generated by SPOT synthesis (Figure 1.14). The resulting membrane was incubated with a His-tagged Tsp solution. The protease Tsp was made inactive by

mutating Ser at 430 to Ala to prevent proteolytic degradation of the immobilized peptides. Following incubation, bound Tsp was identified by the use of an anti-His antibody. This screen identified 24 candidate proteins that potentially interact with Tsp.[38]

Distefano's group adopted a similar approach to make peptide libraries with free C-termini by SPOT synthesis (Figure 1.15).[39,40] In their synthetic scheme, the HMPA-Glu-ODmab in structure **1-3** is the special handle shown in Figure 1.12. The HMPA is the FG 1 and the carboxylic acid side chain of Glu is the FG 2 which is protected by Dmab group. Peptide synthesis continues from HMPA. The N-terminal amino group of the peptide library is protected by Dde group which can be removed along with Dmab group at the same time by hydrazine. The N-terminal amino group and the internal carboxylic acid group are exposed. An intramolecular amide bond formation results in a cyclic peptide. The peptide libraries with free C-termini can be generated by TFA treatment because of the lability of the ether bond between the HMPA and the amino acid X. Note that a photocleavable molecule is incorporated into the peptide synthesis. The purpose of the photocleavable molecule is to release peptides from the membrane by photolysis and be analyzed by MALDI to make sure the library synthesis works fine. They made 20X19 RAGCVa_2X and RAGCCa_2X libraries with the X being one of the 20 proteogenic amino acids except Pro and a_2

being one of the 20 proteogenic amino acids.

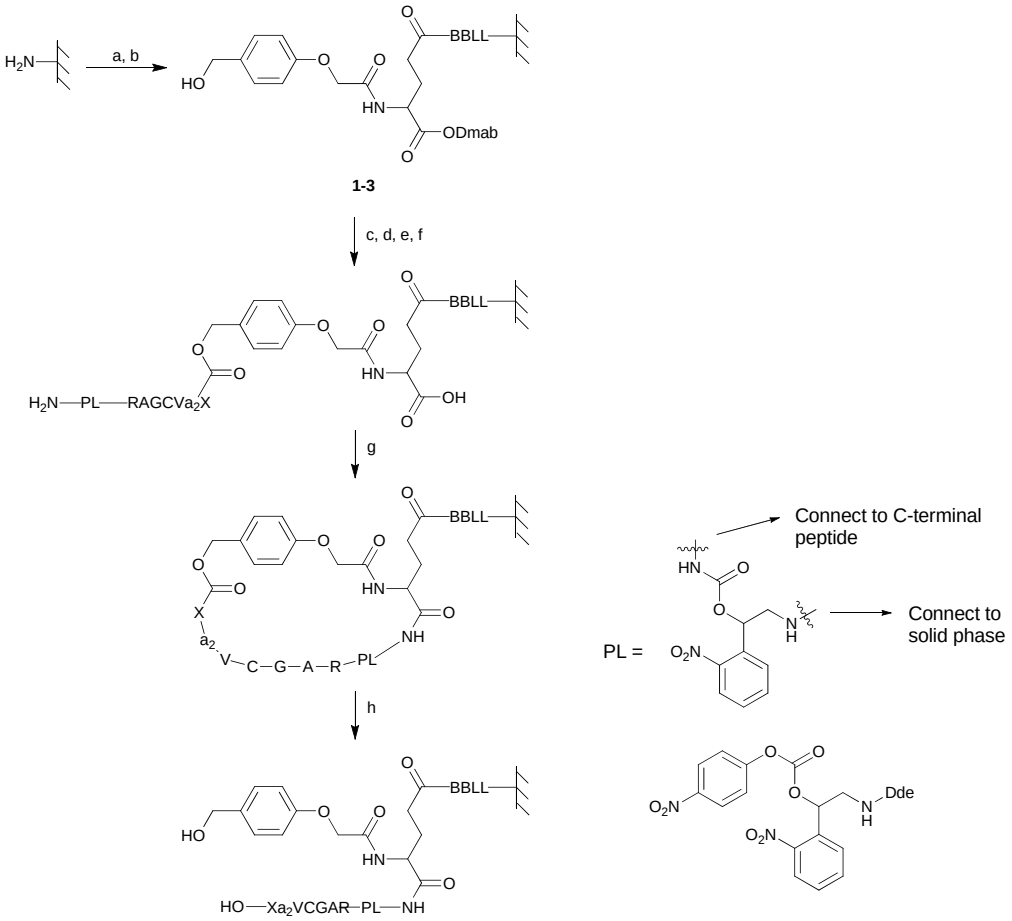


Figure 1.15 Synthesis of inverted peptides for screening assay. Reagents and conditions: (i) standard DIC coupling, then, capping, then 20% piperidine; (ii) standard DIC coupling; (iii) 0.4 M Fmoc-AA and 1.2 M CDI in DMF, then, capping, then 20% piperidine; (iv) standard DIC coupling, then, capping, then 20% piperidine; (v) 0.5 M photocleavable linker, 0.5 M Et_3N in DMF; (vi) 2% NH_2NH_2; (vii) 0.05 M BOP, 0.05 M 6-Cl-HOBt and 0.1 M DIEA in DMF; (viii) modified reagent K.

Distefano's group screened those C-terminal libraries to determine the peptide specificity of protein farnesyltransferase (PFTase). PFTase catalyzes the transfer of a farnesyl (C15) group of farnesyl diphosphate to the C-terminal Ca_1a_2X motif of a

peptide or protein (Figure 1.8). To determine the peptide substrate specificity of PFTase, each membrane was subjected to PFTase-catalyzed prenylation with an alkyne-containing FPP analogue followed by derivatization with biotin-azide via copper-catalyzed azide-alkyne cycloaddition (CuAAC). Those peptides that were prenylated by PFTase were conjugated to biotin at this step. The membrane was then subjected to an enzyme-linked assay involving streptavidin-alkaline phosphatase (SA-AP) and the chromogenic substrate 5-bromo-4-chloro-3-indolyl phosphate (BCIP). Spots containing prenylated peptides appear turquoise colored, whereas spots where the prenylation reaction was inefficient remain colorless (Figure 1.16). They screened the C-terminal libraries to study the specificity of *R. norvegicus* PFTase (rPFTase). They also investigated the interplay between peptides and isoprenoid substrates of varying length (C5, C10 and C15) and the specificity of PFTases from different organisms (*R. norvegicus, S. cerevisiae* and *C. albicans*).[39,40]

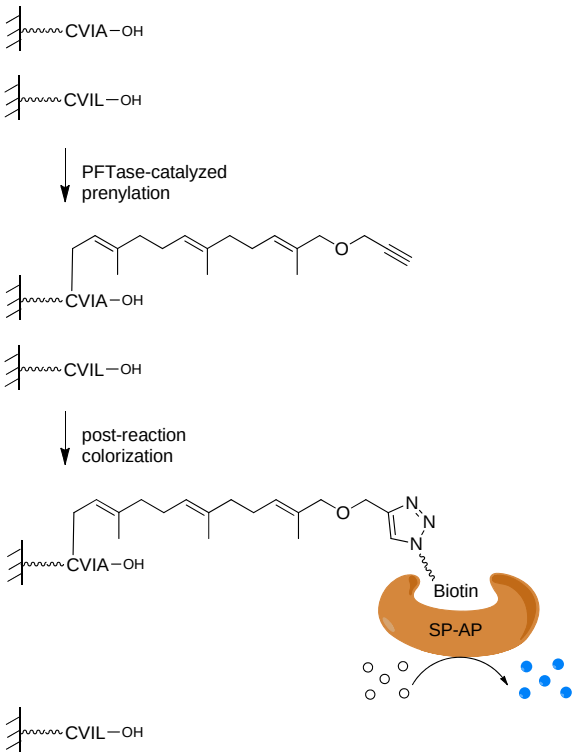


Figure 1.16 Screening and imaging strategy using CVIA (a substrate) and CVIL (nonsubstrate) as examples. Post-reaction colorization was accomplished by click reaction with biotin-azide followed by incubation with SP-AP. Colorization occurred upon the addition of BCIP.

The limitation of the SPOT synthesis is the small library size, usually several hundred members per 10×15 cm^2 membrane. This is a common problem for parallel synthesis. If OBOC method is adopted, it is routine to generate and screen a library of 10^5-10^7 members. However, since all beads are mixed during screening, to determine the sequence of a positive bead (decoding), a tagging procedure should be used during the library synthesis. For example, Bradley and coworkers made an inverted tripeptide library to be screened by guanidinium-based tweezer receptor (Figure 1.17).[41] In their synthetic scheme, 80% of the functional groups of a bead are used to make an

inverted peptide library and the remaining 20% are to incorporate a coding tag. After the C-terminal tripeptide library was generated, it was screened by binding to a solution of dansyl-labeled peptide tweezer receptor or dansyl-labeled peptoid tweezer receptor (Figure 1.18). Highly fluorescent beads were picked up and treated with 100% TFA to expose N-terminal peptides whose sequences are the same as the C-terminal peptides of the same beads. The peptide sequences on those beads can then be sequenced by Edman degradation.

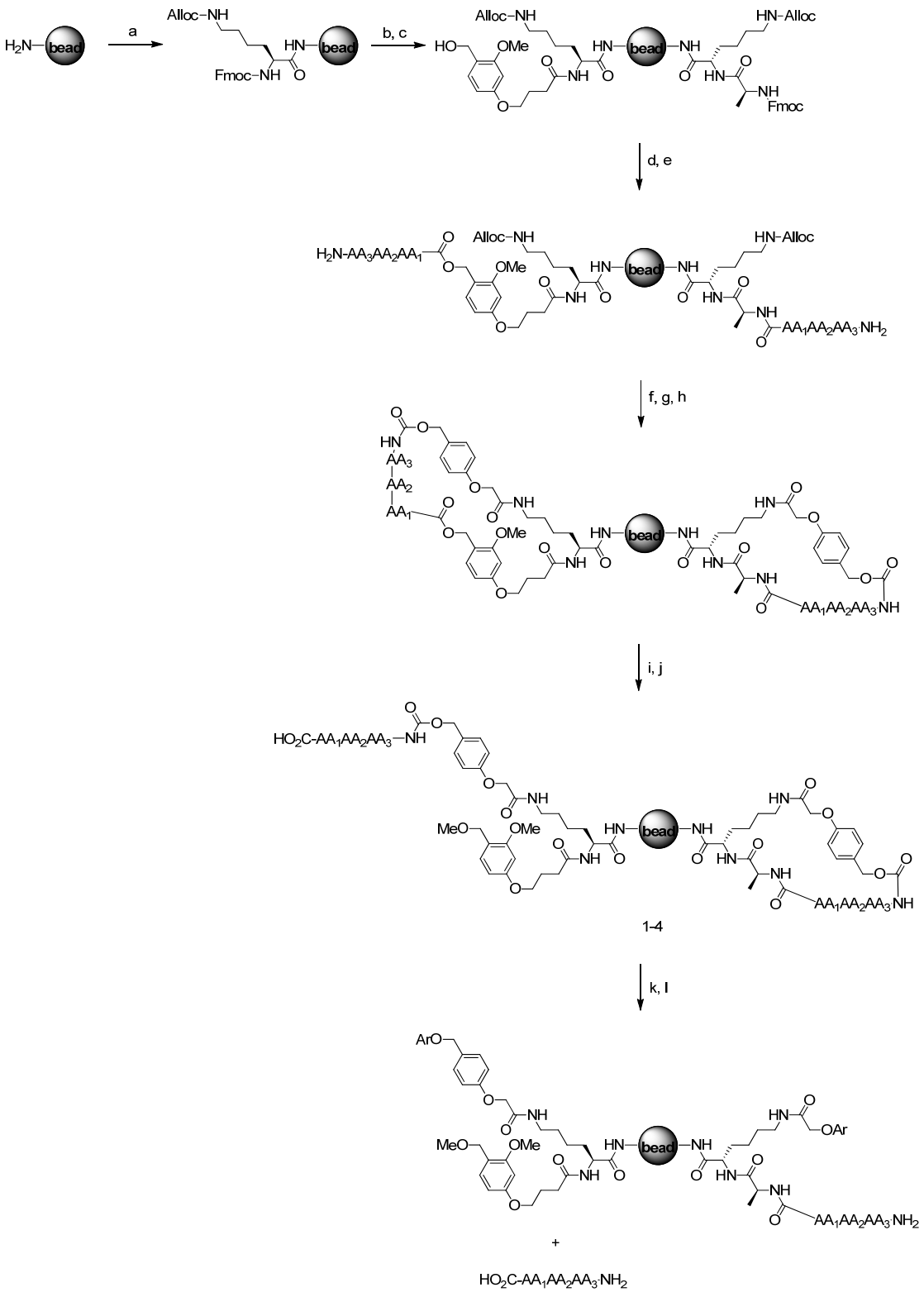


Figure 1.17 Synthesis of resin-bound peptides with free C-termini and post-screening treatment. Reagents and conditions: (a) standard SPPS; (b) HMPB, DIC, DMAP (80% of amine sites); (c) Fmoc-Ala, DIC, HOBt; (d) 20% piperidine; (e) split and mix Fmoc-SPPS; (f)

p-$NO_2C_6H_4OCO_2CH_2C_6H_4OCH_2CO_2$Allyl; (g) $Pd(PPh_3)_4$, dimedone; (h) PyBOP, DIEA, DMAP; (i) 1% TFA / DCM; (j) MeOH; (k) screen linrary (l) treat positive beads with 100% TFA.

(a) (b)

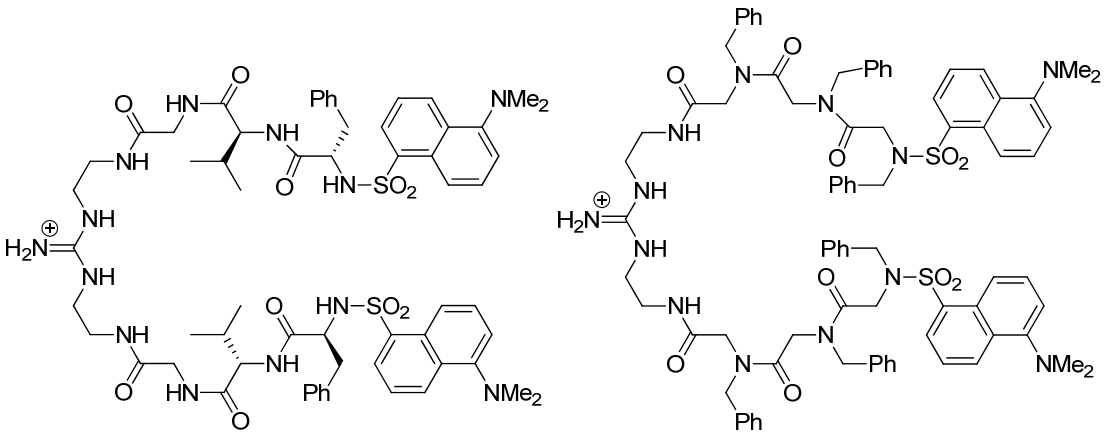


Figure 1.18 Structure of (a) dansyl-labeled peptide tweezer receptor. (b) dansyl-labeled peptoid tweezer receptor.

Saving a portion of the loading of a bead for tagging is a good way for later structural determination. However, if both the inverted peptides and the encoding peptides are randomly attached on the beads as in the above example, the encoding peptides could potentially interfere with library screening. To solve this problem, Lam's group has developed a simple method to spatially segregate TenaGel beads into inner and outer layers.[42,43] TentaGel resins are grafted coplymers consisting of polystyrene matrix on which polyethylene glycol is grafted. TentaGel beads are first soaked in water, drained and then quickly suspended in a mixture of dichloromethane and diethyl ether containing the derivatizing reagent. Because the inner phase of the

bead is still full of water which is not immiscible with the organic solvent, only peptides on the bead surface are exposed to and react with the reagent. The beads are then washed with DMF, and the inner remaining free N-terminal amino groups can react with a different reagent. Pei's group adopted this method to make peptide libraries with free C-termini on the outer layer and N-termini in the inner layer (Figure 1.19).[11,44] About 33% peptides of a bead were used for making a C-terminal peptide library. The majority of the loadings of a bead were used for coding. After making the cyclic structure, the ester bond can be selectively hydrolyzed in $NaOH_{(aq)}$ to generate a peptide with free C-terminus. An inverted peptide library containing five random residues was screened for binding to biotinylated PDZ domains. PDZ-bound resins were identified by subsequent treatment of SP-AP and BCIP buffer. The peptide sequences on those beads were sequenced by partial Edman degradation-mass spectrometry (PED/MS)[5,45] because PED/MS sequencing is less time consuming than traditional Edman degradation sequencing. Pei's group determined the binding specificity of the four PDZ domains of sodium-hydrogen exchanger regulatory factor-1 (NHERF1) and channel-interacting PDZ domain protein (CIPP),[44] and later PDZ domains of the T-cell lymphoma invasion and metastasis 1 (Tiam1) and Tiam2 proteins.[11] They also used a similar strategy to determine the peptide specificity of BUZ domains of histone deacetylase 6 (HDAC6) and mutant ubiquitin processing

protease (Ubp-M) to see if these BUZ domains can bind to other peptide sequences other than Ub.[46] In this work, they used HMPA instead of HMBA as the linker to make the ester bond. Because ether bond between the HMPA and the connected amino acid can be cleaved in acidic condition, ring cleavage and side-chain protecting group removal can be performed at the same time by TFA treatment.

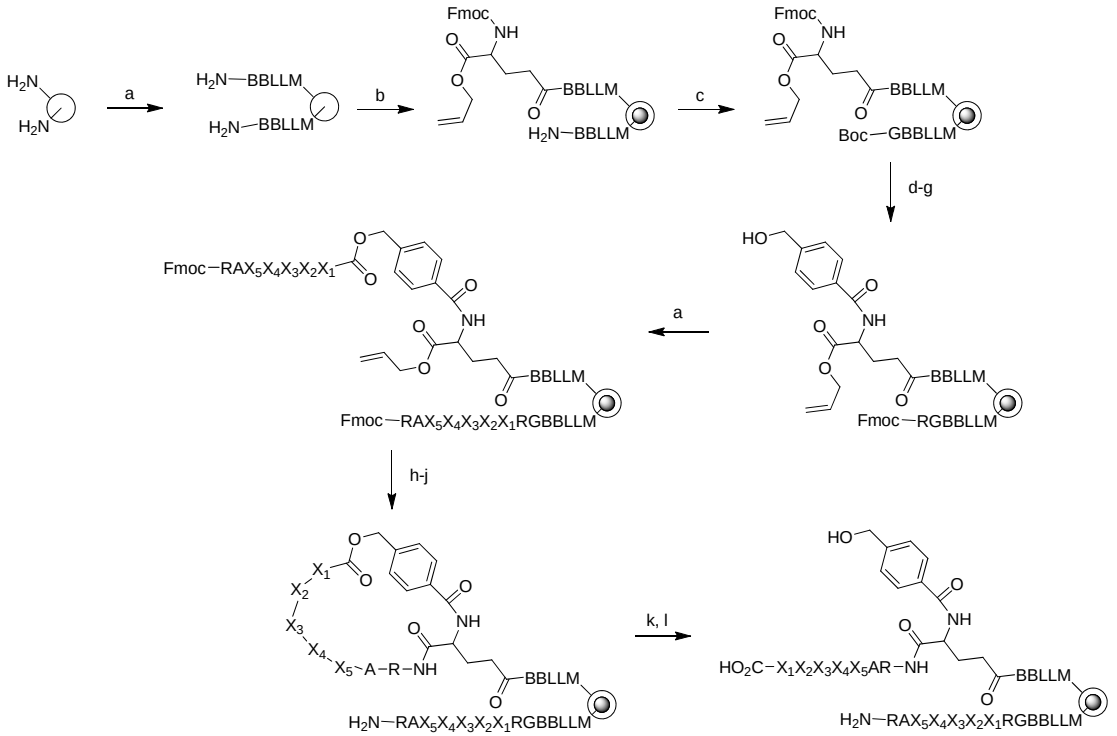


Figure 1.19 Synthesis of spatially segregated and inverted peptide library. Reagents and conditions: (a) standard Fmoc-SPPS; (b) soak in water and then 0.33 eauiv. of Fmoc-Glu(δ-NHS)-OAll in CH_2Cl_2 / Et_2O; (c) Boc-Gly-OH, HBTU; (d) 20% piperidine; (e) HMBA, HBTU; (f) TFA; (f) Fmoc-Arg(Pbf)-OH, HBTU; (h) $Pd(PPh_3)_4$; (i) 20% piperidine; (j) PyBOP, HOBt; (k) $NaOH_{(aq)}$; (l) TFA.

Lam's group used the same bilayer, peptide inversion strategy to discover specific peptide ligand against vancomycin.[47] Vancomycin was labeled with biotin to

facilitate the identification of bound beads. XXXK(Ac)aa tripeptide libraries were screened first (a stands for d-Ala) (Figure 1.20). Later, more focused secondary libraries (YELK(XXX)aa, and XXXXaa) were designed and screened. A ligand with 50-fold higher binding affinity to vancomycin than the native ligand Kaa was identified.

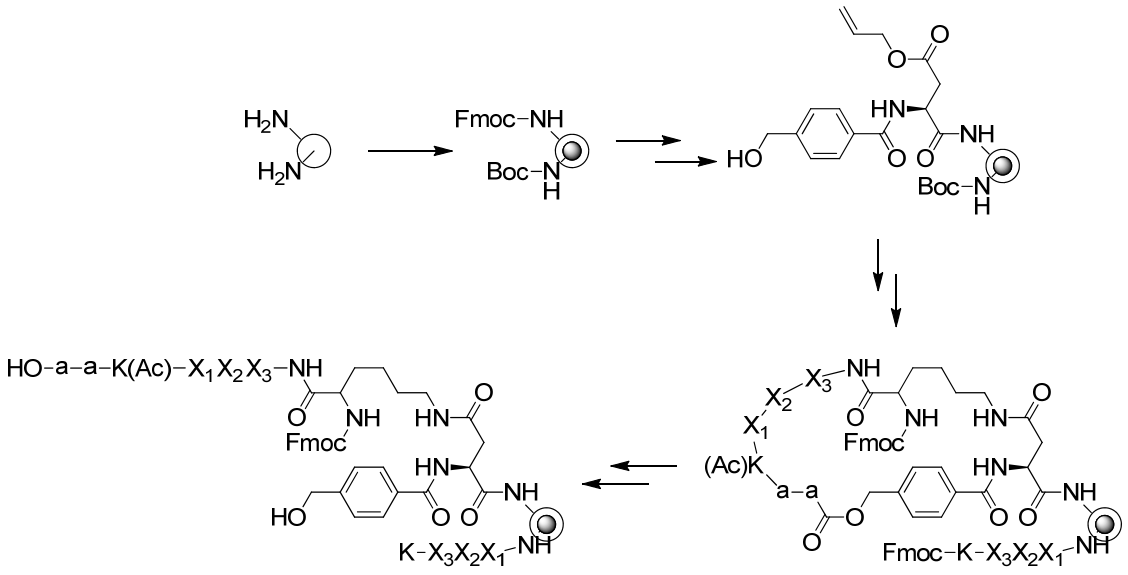


Figure 1.20 Synthesis of spatially segregated and inverted peptide library.

(c) Phage display

As mentioned in part b of section 1.1, many types of phages have been adapted for phage display. If phage display is used to screen C-terminal peptide libraries, one must choose a phage and a coat protein of it that has C-terminus exposed so that a

peptide library can be appended to its C-terminus. λ phage is a good choice (Figure 1.21).[48] The carboxyl terminus of the D-capsid protein does not involve in head formation[49] so random peptides can be displayed on the surface of λ phage by fusion to the carboxyl terminus of D proteins. After the DNA library of λ display vector is prepared, it can be incorporated into phage particles by in vitro packaging. Those λ phages can be propagated by infecting the host bacteria cells. After the phage multiply, their progeny will display the corresponding protein sequence on their surface along with the head proteins.

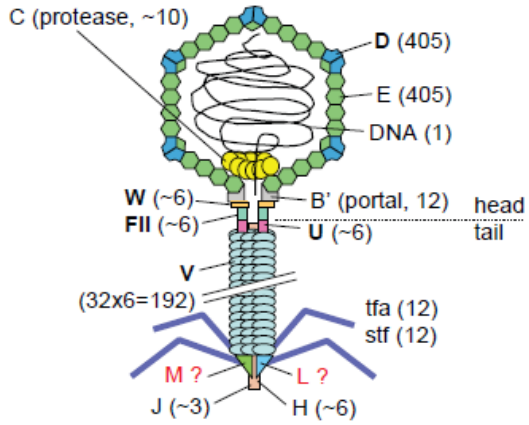


Figure 1.21 A model of phage λ, indicating protein-protein interactions. Proteins in bold font have known structures. Numbers indicate the number of protein copies in the particle.[48]

Dente and coworkers used λ phage to display nonapeptides of random sequences with free C-termini.[9] In their system, about 95% of the total D protein is chimeric and the remaining 5% is synthesized by a second wild-type gene in the phage vector. Seven PDZ domains of the human INADL protein were expressed, immobilized on

beads and screened to determine their binding specificities. Different consensus binding sequences were defined for each hINADL PDZ domain. Later they used a similar λ phage strategy to determine the five PDZ domains in the protein tyrosine phosphatase PTP-BL.[50] In this work they not only screen and find the specificity of each separate PDZ domain but also investigate the binding preferences of several combinations of PDZ domains. Interestingly while a protein segment spanning the last two PDZ domains of PTP-BL (PDZ4+5) did not reveal an altered binding preference relative to PDZ4, the combination of the first two PDZ domains (PDZ1+2) displayed a more stringent selectivity than PDZ2 alone. This demonstrates that there is an allosteric interaction between PDZ1 and PDZ2, and the ligand selectivity of PDZ2 can be modulated by PDZ1.

T7 phage is also a good choice for C-terminal peptide display. The procedure of T7 phage display is similar to λ phage display except the peptide-coding sequence to be displayed is cloned into the phage DNA to make a translational fusion to the head protein, gp10, of the phage (Figure 1.22).[51] Since there are 415 of head proteins per phage, a large number of copies of peptide can also be expressed from each phage, enhancing binding. Bredesen and coworkers used T7 phage display to determine the specificity of six PDZ domains from the synaptic proteins PSD-95 and SAP97.[52] Later Spaller and coworkers also used T7 display method to determine the specificity

of the 10th PDZ domain of the multi-PDZ domain protein 1 (MUPP1) and the 3rd PDZ domain of PSD-95.[53]

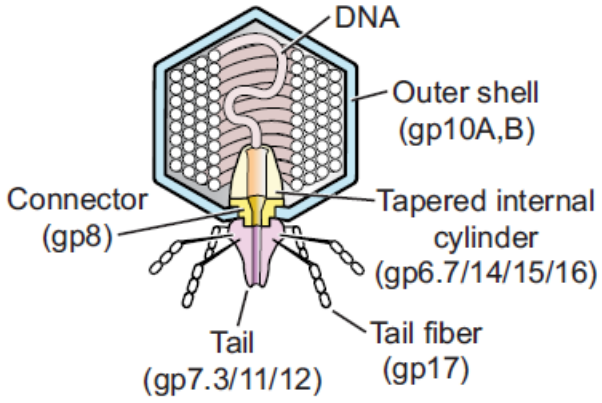


Figure 1.22 The structure of bacteriophage T7.[51]

Ruoslahti and coworkers used T7 phage display to identified peptides that can bind to human prostate cancer cells PPC-1.[54] The cancer cells were incubated with the phage libraries, washed, and then lysed to release the phages. The selected phages were determined to display a consensus sequence R/K-X-X-R/K. The strict requirement for a basic C-terminal amino acid is termed C-end rule. The bound phage could also be internalized into the cells at 37 ^{o}C. Affinity chromatography and immunostaining identified cell receptor NRP-1 as the target responsible for the C-end rule. NRP-1 is a known mediator of vascular permeability.[14,15] Intravenously injected positive phages penetrated into tissue parenchyma, but not negative control, indicates that the C-end rule peptides can cause vascular leakage and tissue penetration.

M13 phage is another common phage used for phage display. In contrast to some phages that lyse the cell to release themselves such as λ, T4 and T7 phages, M13 phage is secreted through the membrane of bactreria. Of the five coat proteins of the M13 phage, pIII and pVIII are the most commonly used coat proteins to make translational fusion for phage display (Fidure 1.23). In the M13 phage display technique, after a DNA library of M13 display vector is prepared, they are transformed into *E. coli* cells. However, the M13 vector alone does not encode all the necessary proteins to make progeny phage. Then the cells are infected with a helper phage which supplies normal head protein. So in the assembled phage, the fusion proteins are usually mixed with the wild-type protein in the head.

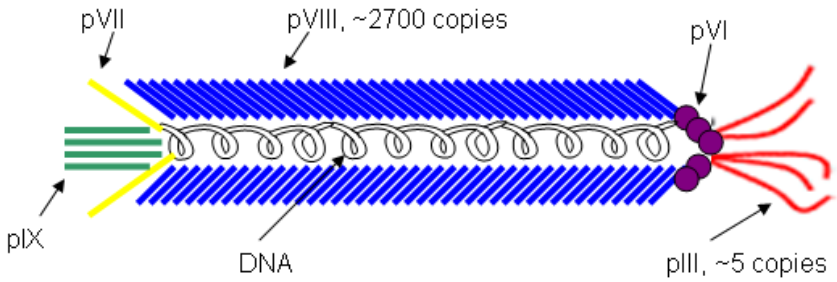


Figure 1.23 Schematic representation of the filamentous bacteriophage M13.

Unfortunately there is a problem in our case: in the natural assembly of M13 phage, it is the N-terminus of pIII and pVIII that is solvent-exposed whereas the carboxyl terminus is buried in the particle.[55,56] So in M13 phage display, the random

peptide sequence is usually inserted between the signal sequence required for secretion of the head protein and the N-terminus of the mature protein where it will be displayed on the surface of the phage. However, Sidhu and coworkers found that although the pVIII C-terminus is buried in the particle core, C-terminal fusions can still be displayed. They screened a heptapeptide library to investigate the binding specificities of PDZ domains 2 and 3 of MAGI 3, a membrane-associated guanylate kinase.[57] In their M13 phage system, there is an intervening linker of 13 residues between the C-terminus of pVIII and the random peptide to make sure the C-terminal peptides can be accessible for interaction to the target protein. They also used this methodology to identify ligands for the Erbin PDZ domain.[58] Later they improved their system by screening libraries of mutated pVIII protein (plus a ten residue linker) for better C-terminal peptide display. Interestingly, the selected pVIII mutants containing mutations spanning near the N-terminus to near the C-terminus. Their optimal pVIII mutants showed C-terminal display more than 100-fold relative to the wild-type.[59] They used the optimized methodology to determine the peptide specificity of about half of the PDZ domains known to exist in the human and *C. elegans* proteomes.[60] They also screened the M13 phage libraries of all human and viral C-terminal peptides against nine PDZ domains of human Densin-180, Erbin, Scribble, and Disks large homolog1.[61]

Yin's group used a different strategy to display protein libraries connecting to pIII of M13 phage. pIII, like pVIII, has its N-terminus exposed to solvents. One clever way to use pIII to display a C-terminal library is to link the peptide library to pIII in the periplasm of the cell after it has been secreted instead of translationally fusing the library to pIII. In their system, pIII and the protein to be displayed are expressed from different genes on the M13 display vector and both have a signal sequence. A Jun peptide sequence is placed before the N-terminus of pIII, and a Fos peptide sequence is placed before the N-terminus of the protein library. The Jun-pIII and the Fos-protein will be secreted separately into the periplasm and form a heterodimer due to the binding of Jun-Fos coiled coil and two disulfide bonds between Jun-Fos. In this way, phage assembly anchors the Jun-Fos dimer on the phage surface and the displayed protein is connected to the dimer with an exposed C-terminus (Figure 1.24). Because the cross-linking of the two proteins occurs before the phage leaves the cell, the vector DNA that encodes both of them is packaged in the phage head.

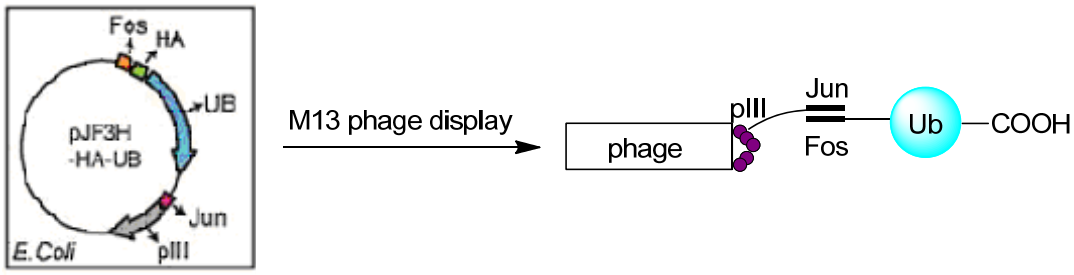


Figure 1.24 Phage display of Ub with free C-terminus.[62]

Yin's group is interested in the ubiquitin modification pathway. Ub and UBLs are activated by specific E1 enzymes (Figure 1.10). It was reported that the C-terminal peptide of Ub plays a key role in recognition of E1 by crystal structure.[63] A Ub library with random sequences from residue 71 to 75 (Ub has 76 residues) was screened to determine the specificity of 2 E1 enzymes (Ube1 and Uba6).[64] The E1 protein was immobilized on a solid support. Selection of phage-displayed Ub library is based on the formation of a Ub-E1 conjugate. The enriched phage particles can be eluted with DTT treatment by cleaving the thioester conjugate (Figure 1.25).[64] Later they used a similar strategy to screen a C-terminal library of Nedd8, a ubiquitin-like protein, to profile the specificity Nedd8 activating E1 enzyme (NAE).[65] They also made a mutant E1 (termed A1) which cannot activate wild-type Ub and screened a Ub library against the A1 by the above screening strategy to find Ub mutants which can be recognized by A1. Subsequent optimization resulted in an orthogonal Ub-E1 system in which the engineered Ub can be transferred to A1 but there is no cross-activities with wild-type Ub and E1.[62]

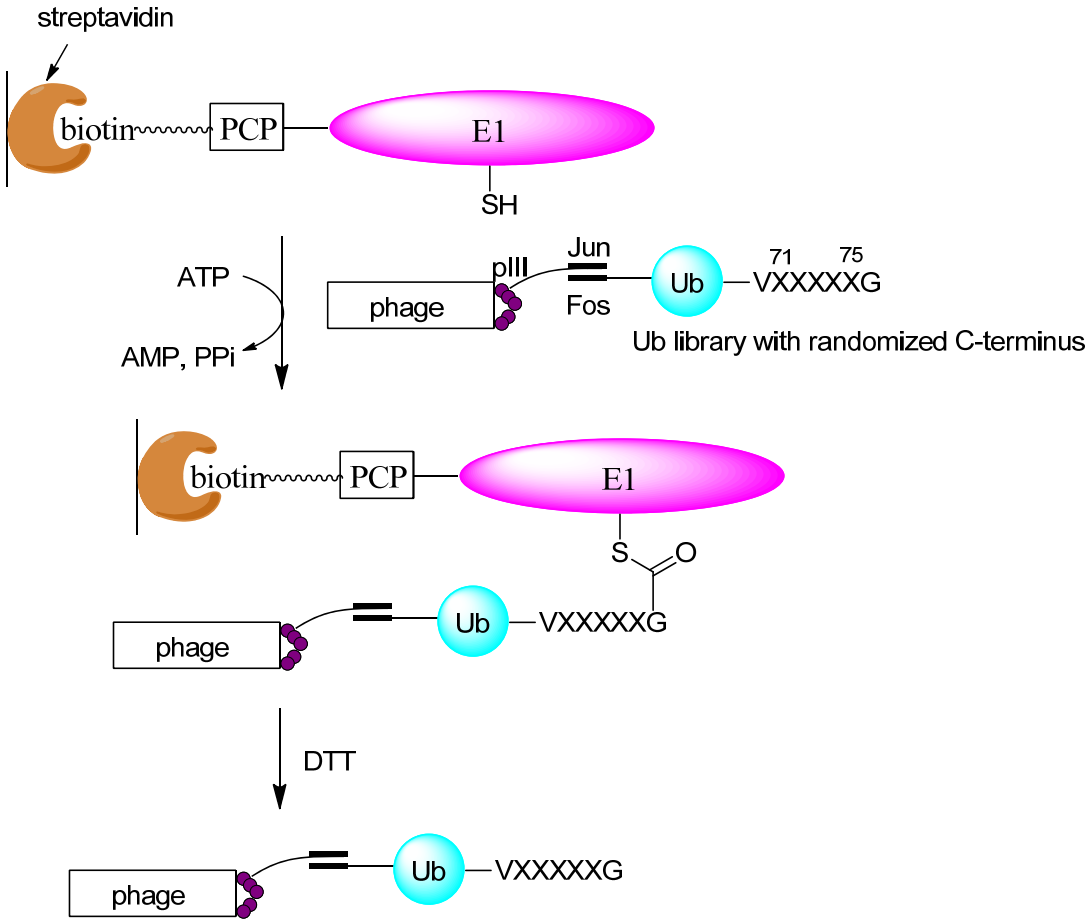


Figure 1.25 Phage selection of the Ub library.[64]

1.3 Summary

A variety of methods to synthesize and screen peptides have been reviewed in this chapter with the focus on peptide libraries with free C-termini. Each method has its own advantages and disadvantages. The principle of SPOT synthesis is simple. Different peptides are displayed on different locations of a membrane. The positions of the spots on the membrane tell the corresponding peptide sequences. This makes the sequence determination easier whereas the beads are mixed in OBOC method

making a tagging procedure needed. However, the library size of SPOT synthesis is really small compared to OBOC method or phage display making the sampling of all chemical space difficult. Phage particles can display protein on its coat protein. Hence, protein libraries can be screened by phage display whereas in solid-phase peptide synthesis, a peptide sequence within 30 residues is usually synthesized linearly. Phage libraries can also be injected into animals to do *in vivo* screening.

Because the solid support of SPOT synthesis and OBOC method is inert, chemical reactions can be performed easily on it whereas proteins tend to denature in organic solvent making phages lose activity. Unnatural amino acids (UAAs) can be easily incorporated into peptides by SPPS but Nature encodes 20 amino acids only. Although enzymatic transformation or genetic code expansion technique are able to diversify peptides made by cellular machinery, the diversity of the resulting peptides still cannot compete with chemical synthesis. A summary of the advantages and disadvantages of those methods is shown below (Table 1.1).

Category	Advantage	Disadvantage
SPOT synthesis	Easy to decode sequences Easy to incorporate UAAs Easy to do chemical derivatization	Small library size (~10^3)
OBOC	Big library size (~10^6) Easy to incorporate UAAs Easy to do chemical derivatization	Tagging is needed to decode sequences Cannot make protein libraries
Phage display	Big library size (~10^9) Can display peptide or protein libraries In vivo screening is possible	Difficult to incorporate UAAs Unstable in organic solvent

Table 1.1 A comparison of SPOT synthesis, one-bead-one-compound and phage display methods for synthesis and screen peptide libraries.

Chapter 2. Isoprenoid Analogues Used in the Study of Protein Prenylation

2.1 Introduction to prenyltransferases

After protein is synthesized, the structure can be modified by other enzymes. For example, a protein can be cleaved by a protease, phosphorylated by a kinase or lipidated by a prenyltransferase. This process is called post-translational modification. The purpose of post-translational modification is to switch on (or switch off) an enzyme's activity, to locate the protein to certain region in the cell or to change the binding affinity to other proteins. Protein prenylation with isoprenoids has been the focus of numerous studies since its discovery in the early 1990's because of its connection to cancer.[66] Members of the Ras family proteins are normally prenylated and mutated forms of Ras, especially K-Ras, are involved in as many as 30% of all human cancers.[67] Members of the protein prenyltransferase class of enzymes include protein farnesyltransferase (PFTase) and geranylgeranyltransferase I (GGTase I). These prenyltransferase enzymes are heterodimers. The α-subunit of PFTase and GGTase-I are the same but the β-subunit of these proteins are only about 25% sequence identical[68] (Figure 2.1).

(a) (b)

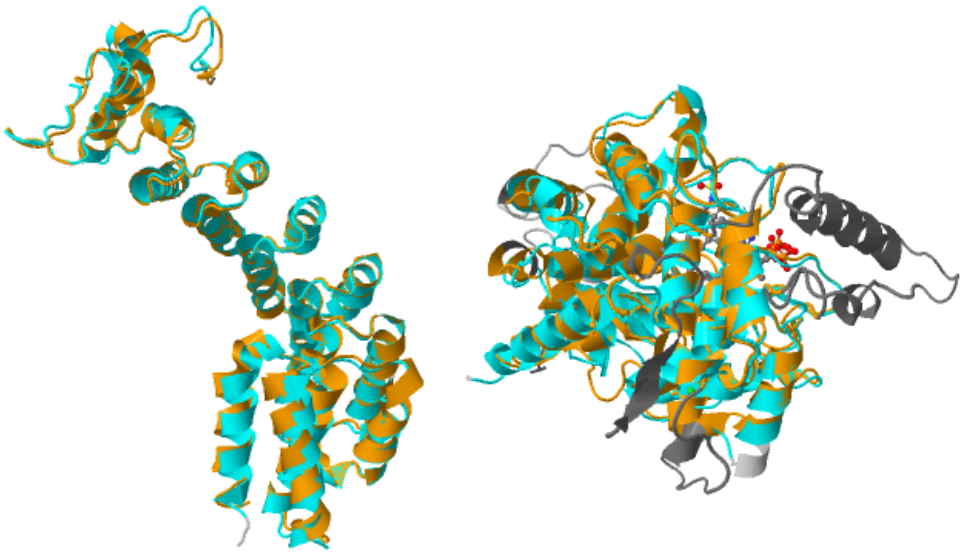


Figure 2.1 Alignment of the crystal structures of PFTase (brown, PDB 2BED) and GGTase-I (cyan, PDB 1N4P). (a) α-subunit. (b) β-subunit. Structures were overlaid using "Compare Structures" function on the RCSB website.

PFTase and GGTase-I catalyze the transfer of farnesyl (C15) and geranylgeranyl (C20) groups respectively, from the corresponding diphosphates. In the resulting alkylated protein, the isoprenoid group is linked to a cysteine residue within the C-terminal amino acid sequence referred to as a Ca_1a_2X motif where C is a cysteine, a_1 and a_2 are usually aliphatic amino acids and X is the major determinant for modification by either PFTase or GGTase I (Figure 2.2).[19,20] It has been reported that farnesylation by PFTase occurs when X is alanine, serine, methionine or glutamine, while geranylgeranylation by GGTase I occurs when X is leucine or phenylalanine. Additional proteins are di-geranylgeranylated at their C-termini when they contain

sequences including CC and CXC; these latter sequences are prenylated by protein geranylgeranyltransferase type II (GGTase-II). GGTase-II, also called RabGGTase, differs both structurally and functionally from the canonical PFTase and GGTase-I because it recognizes more extensive elements from its cognate protein substrates, Rab proteins.[69] Once prenylated, the resulting proteins move to the endoplasmic reticulum where they are further processed by proteases that remove the a_1a_2X tripeptide and methylated by a SAM-dependent methyltransferase to produce proteins with a C-terminal methyl ester.[70]

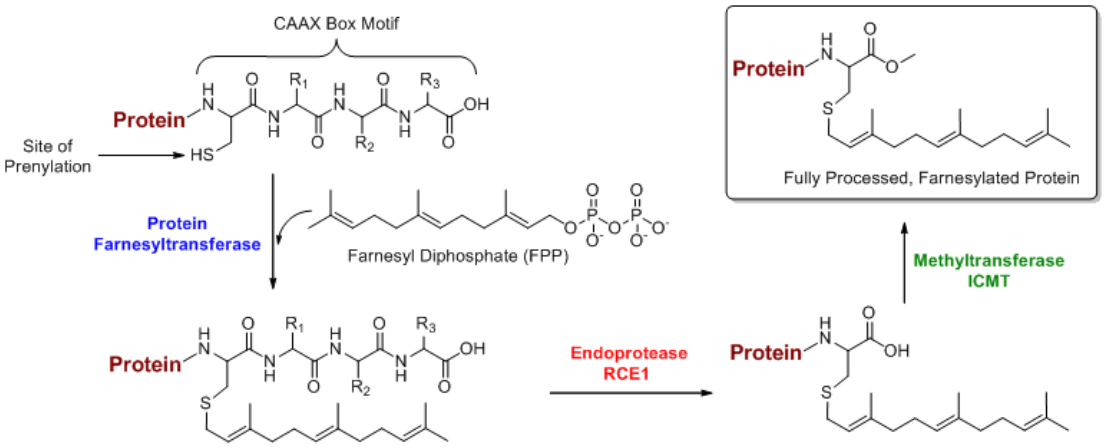


Figure 2.2 C-terminal prenylation and subsequent proteolysis and methylation. The prenyl group could be farnesyl group (C_{15}) (shown in this figure) or geranylgeranyl group (C_{20}).

Protein prenylation is not only common in mammals[71,72] but is also a ubiquitous post-translational modification in all eukaryotes. For example, prenylated Ras is a more potent in vitro activator of *S. cerevisiae* adenylyl cyclase than is the

non-prenylated form.[73] It has also been found that prenylation of signal transduction proteins is essential for viability of *C. albicans* which is an opportunistic fungal pathogen.[74] Other groups have identified prenylated proteins and confirmed their significance in *P. falciparum* which is the causative agent for malarial disease.[75] A vast array of prenylation inhibitors have been developed to combat numerous illnesses caused by cancers, protozoan pathogens, and fungal infections.[76-79] Recently, interest in prenylation has expanded to include biotechnology applications since prenyltransferases can be used to enzymatically incorporate non-natural functional groups into protein substrates.[80-82] The resulting modified polypeptides can be further transformed via bio-orthogonal reactions to produce a variety of useful species including PEGylated proteins,[83] protein multimers[84] and protein-DNA conjugates.[85]

2.2 Fluorescent analogues

Since the native isoprenoid substrates of prenylation (FPP and GGPP) do not have special spectroscopic properties, this makes the monitoring the localization of intracellular proteins and biochemical characterizations difficult. The fluorescent derivatives (Figure 2.3) can be used to label proteins *in vitro* or in live cells and have the application for studying the interaction and subcellular localization of prenylated proteins.

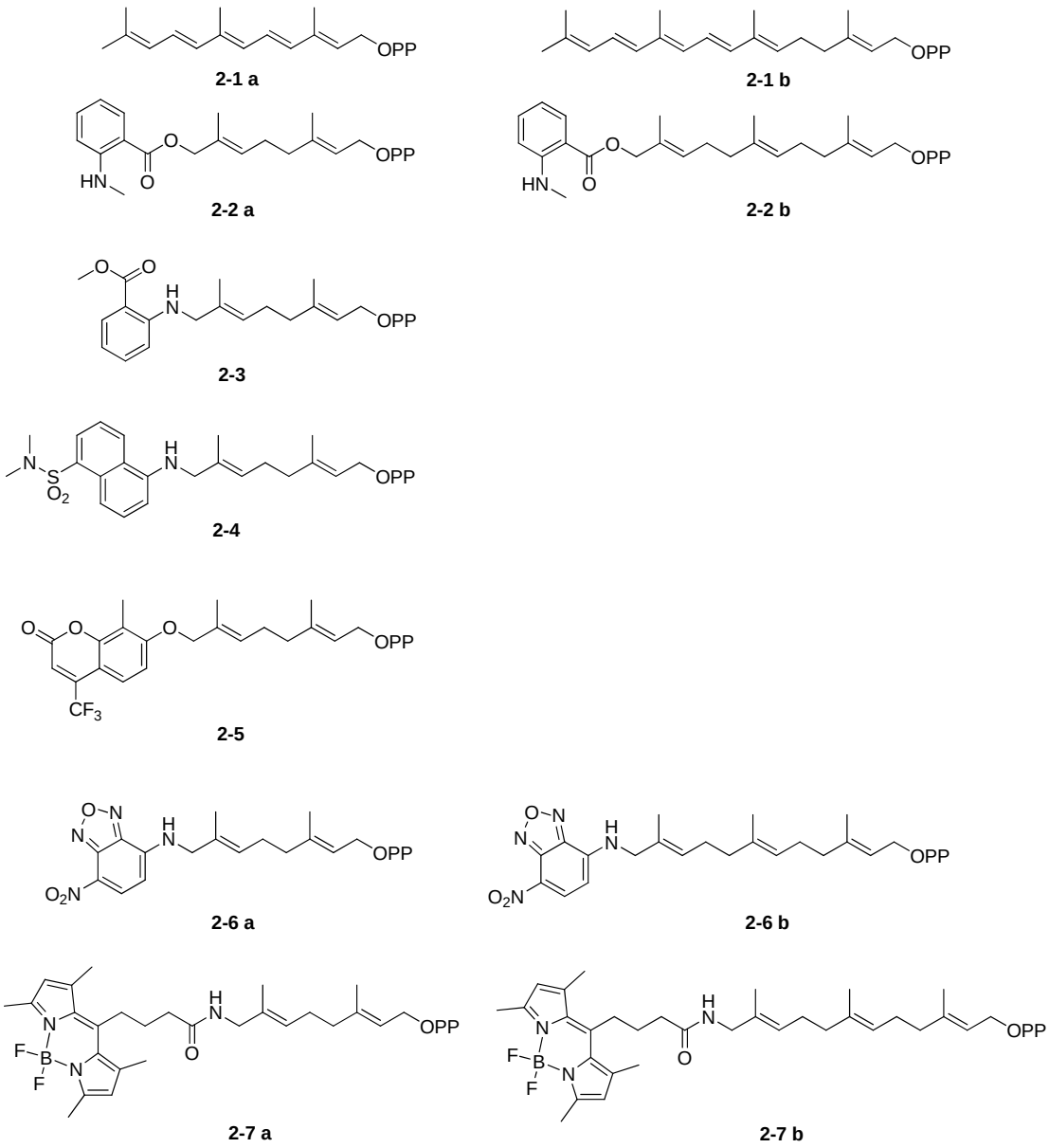


Figure 2.3 Fluorescent analogues of FPP or GGPP.

The first way to derivatize FPP or GGPP into fluorescent molecules is to extend double bond conjugation to make FPP analogue (ΔΔFPP, **2-1 a**) and GGPP analogue (ΔΔGGPP, **2-1 b**).[86,87] The alcohol of **2-1 b** (ΔΔGGOH) showed blue fluorescence in methanol, with λ_{ex} = 310 nm and λ_{em} = 410 nm but the fluorescence intensity at 410

nm was largely quenched in aqueous solution. Side-chain prenylation of N-acetylated Cys methyl ester with **2-1 b (**AcCys(ΔΔGG)Me) showed λ_{ex} = 360 nm and λ_{ex} = 460 nm in Tris buffer demonstrating the environmental sensitivity of the ΔΔGG fluorophore. **2-1 b** is a substrate for yeast GGTase-I with a K_M value of 0.33 μM (3.45 μM for GGPP) but the prenyl-transfer efficiency was lower than that of GGPP. The dissociating constant (K_D) of AcCys(ΔΔGG)Me to RhoGDI, which forms a complex with geranylgeranylated Rho protein, is 15.1 μM (2.45 μM for AcCys(GG)Me) indicating that the more rigid ΔΔGG had lower affinity. The alcohol of **2-1 a** (ΔΔFOH) displayed blue fluorescence in methanol, with λ_{ex} = 360 nm and λ_{em} = 465 nm but again the fluorescence is quenched with increasing content of water. Unfortunately **2-1 a** is not a substrate for yeast PFTase but a potent competitive inhibitor (Table 1).

All of the subsequent fluorescent analogues are derived from appending an extra fluorophore onto the isoprenyl chain. The first two examples of such analogues are N-methylanthranilate-labeled FPP analogue (mGPP, **2-2 a**) and GGPP analogue (mFPP, **2-2 b**).[88,89] Both compounds had λ_{ex} at 340 nm and λ_{em} at 435 nm and both compounds are substrates for GGTase-II with **2-2 b** binding better than **2-2 a** because **2-2 b** is structurally more similar to GGPP.

The first advantage of appending a fluorophore onto the isoprenoid chain is to determine binding affinity and kinetics. Previous attempts to determine these

properties depend on changes in the intrinsic tryptophan fluorescence in the active site but in some cases the intensity changes are small. If an extra fluorophore is attached to the isoprenoid chain and binding of the fluorescent derivatives to the enzymes results in significant fluorescence changes, binding affinity can be measured by fluorescence titration. Then the binding affinity of non-fluorescent substrates can be determined by competitive fluorescence titration. In fact, the binding constants of **2-2 a** and **2-2 b** for GGTase-II were determined in this way. The transient kinetics of **2-2 a**, **2-2 b**, FPP and GGPP interaction with GGTase-II were determined similarly but the signal was from fluorescence resonance energy trabsfer between tryptophan and N-methylanthranilate.

Although compound **2-2 a** and **2-2 b** have applicability on enzymatic assays, the ester linkage in the structure might be cleaved by nonspecific esterase in cellular experiments. A more stable anthranylate derivative with amine linkage (**2-3**) should overcome this problem. Compound **2-3** has λ_{em} at 435 nm. Incubation of cells with compound **2-3** led to the accumulation of non-prenylated Ras, which can be prevented by FPP, indicating compound **2-3** is likely a competitive inhibitor with respect to FPP.[90] The alcohol version of compound **2-3** is toxic to RPMI-8402 human leukemia cells with IC_{50} of ~10 nM. The fluorescent analogues **2-4** and its alcohol version could also be visualized within cells by fluorescent microscopy but their other properties

were not further demonstrated.

Compound **2-5** was originally designed to study Undecaprenyl pyrophosphate synthesase (UPPs) which produces C_{55} undecaprenyl pyrophosphate pyrophosphate (UPP) via consecutive condensation reactions of eight IPP with an FPP.[91] In aqueous solution, compound **2-5** displays λ_{ex} = 336 nm and λ_{em} = 460 nm. Due to the electron-withdrawing ability of CF_3, the lowest unoccupied molecular orbital (LUMO) is lowered compared to the CH_3 counterpart resulting in red shift of both λ_{ex} and λ_{em}. Compound **2-5** is not only an inhibitor for UPPs but also a slow substrate for UPPs. The K_M of compound **2-5** is consistent with its K_i and is also similar to the K_M of FPP but the reaction is 125-fold slower compared to the reaction using FPP. Because UPPs can quench the fluorescence of compound **2-5**, the binding process of UPPs was monitored by stopped-flow spectrofluorimeter to calculate transient kinetics. Distefano's lab also found that compound **2-5** is a slow substrate for PFTase because of structural bulkiness of compound **2-5** relative to FPP. The molecular volume of compound **2-5** is about 25% bigger than FPP. k_{cat}/K_M of compound **2-5** is about 380 times lower than FPP. However, it was shown that if the active site of PFTase is expanded by mutating Tyr at 205 to Ala, the mutant enzyme can use compound **2-5** much more efficiently (k_{cat}/K_M increases about 300 times compared to wild-type enzyme) (Table 1).

In the design of fluorescent isoprenoid analogues, the fluorophore should be small to mimic the natural substrate but the fluorescent intensity should also be high to make the measurement easy. 7-nitro-benzo[1,2,5]oxadiazol-4-ylamino (NBD) is one of the fluorophore that fulfills the compensation between structure minimum and fluorescent intensity maximum. FPP analogues (NBD-GPP, **2-6 a**) and GGPP analogues (NBD-FPP, **2-6 b**) have λ_{ex} = 487 nm and λ_{em} = 550 nm.[92] Addition of excess PFTase, GGTase-I and GGTase-II to a solution of **2-6 a** or **2-6 b** resulted in large fluorescent changes. This allowed the binding constant to be determined by fluorescence titration (Table 1).

The second advantage of appending a fluorophore onto the isoprenoid chain is visualizing prenylated proteins easily if that isoprenoid analogue is a substrate. In the in vitro prenylation assay, prenylated protein can be resolved on SDS-PAGE and fluorescence intensity be quantified as exemplified with **2-2 a**, **2-2 b**, **2-6 a** and **2-6 b**. If the sensitivity of the assay is high enough, steady-state kinetics can be measured as in the case of **2-6 a** and **2-6 b** (Table 1). In the cell labeling, in vivo conjugation can be imaged if the isoprenoid analogue is stable. Transiently expressed EYFP-Ki-Ras and ECFP-RabGDI in COS-7 cells in the presence of **2-6 a** or **2-6 b** can be seen on subsequent SDS-PAGE analysis. Intracellular distribution of compound **2-6 b** in human epidermal carcinoma A431 cells was also seen under fluorescence

microscope.[92]

BODIPY-labeled derivatives **2-7 a** and **2-7 b** displayed large fluorescence changes upon binding to protein prenyltransferase, which allowed for the determination of their affinities using fluorescence titration experiments.[93] The results showed that **2-7 a**, **2-7 b**, and FPP have similar affinities for PFTase and **2-7 a** and GGPP have similar affinities for GGTase-I. However, GGTase-II displayed an affinity for **2-7 a** approximately an order of magnitude lower compared to FPP. Unfortunately among the three enzymes, only GGTase-II can utilize **2-7 a** to prenylate its protein substrates. PFTase and GGTase-I could not use **2-7 a** as a substrate. Time-resolved changes in polarization of BODIPY fluorophore was used to determine K_M for GGTase-II-mediated transfer of **2-7 a** onto Rab 7, suggesting **2-7 a** is an efficient substrate for GGTase-II. Compound **2-7 a** was also used to monitor its localization in baby hamster kidney (BHK) cells and in zebrafish.

Compound	Fluorescence property	Substrate? Inhibitor?	Kinetic constants
FPP	ND	S	yPFTase: K_m = 27 μM (DsGCVLS)[87] mPFTase: K_m = 46 nM[94], 28 nM[95] or 1.5 uM[96] (DsGCVLS) mGGTase-II: K_d = 60 nM[89]
GGPP	ND	S	yGGTase-I: K_M = 3.45 μM[86] (DsGCVLL) mGGTase-I: K_d = 0.25 nM[92] mGGTase-II: K_d = 8 nM[89]
2-1 a	λ_{ex} = 360 nm, λ_{rm} = 465 nm in methanol	I	yPFTase: Ki = 8.8 μM[87]
2-1 b	λ_{ex} = 310 nm, λ_{rm} = 410 nm in methanol	S	yGGTase-I: K_M = 0.33 μM[86] (DsGCVLL)
2-2 a	λ_{ex} = 340 nm λ_{rm} = 435 nm	S	mGGTase-II: K_d = 330 nM[89]
2-2 b		S	mGGTase-II: K_d = 49 nM[89]
2-3	λ_{ex} = ~340 nm λ_{rm} = 422 nm	I	mPFTase: IC_{50} = ~ 50 nM[90]
2-5	λ_{ex} = 336 nm λ_{rm} = 460 nm	S	mPFTase (WT): K_M =1.4 μM, k_{cat} =1.3×10^{-3} s^{-1} (GfP-CVIA) [a] mPFTase (Y205A): K_M =0.12 μM, k_{cat} =0.035 s^{-1} (GfP-CVIA) [a]
2-6 a	λ_{ex} = 487 nm λ_{rm} = 550 nm	S	mPFTase: K_d = 1.6 nM K_M = 0.5 μM, k_{cat} = 0.09 s^{-1} (K-Ras)[92]
2-6 b		S	mGGTase-I: K_d = 6 nM K_M = 0.03 μM, k_{cat} = 0.08 s^{-1}

			(RhoA)[92] mGGTase-II: K_d = 328 nM K_M = 4.7 μM, k_{cat} = 0.04 s^{-1} (Rab7)[92]
2-7 a	λ_{ex} = 490 nm λ_{rm} = 505 nm	S only for GGTase-II	mPFTase: K_d = 1.1 nM mGGTase-I: K_d = 0.23 nM mGGTase-II: K_d = 245 nM; K_M = 0.2 μM (Rab7)[93]
2-7 b		I	mPFTase: K_d = 4.9 nM

[a] Dozier, J. K.; Khatwani, S. L.; Wollack, J.; Wang, Y.; Schmidt-Dannert, C.; Distefano, M. D. *Bioconjugate Chem.* **2014**, DOI: bc-2014-00240p.R1

Table 2.1. Summary of fluorescent isoprenoid analogues

2.3 Probes to detect prenylome

Protein prenylation is a widespread phenomenon in eukaryotic cells that affects many important signaling molecules. However, the total number of prenylated proteins which are referred to prenylome is still not clear. Only a fraction of predicted prenylated proteins were experimentally confirmed, leaving the true size of the eukaryotic prenylome unknown. Prenylated proteins have been traditionally detected by metabolic labeling with radioactive substrates such as [^{3}H]FPP and [^{3}H]GGPP. However, this method suffers from some disadvantages such as low sensitivity and long exposure times for detection (weeks to months). To address these limitations, chemical probes of protein prenylation have been developed to improve the detection of prenylated proteins (Figure 2.4).

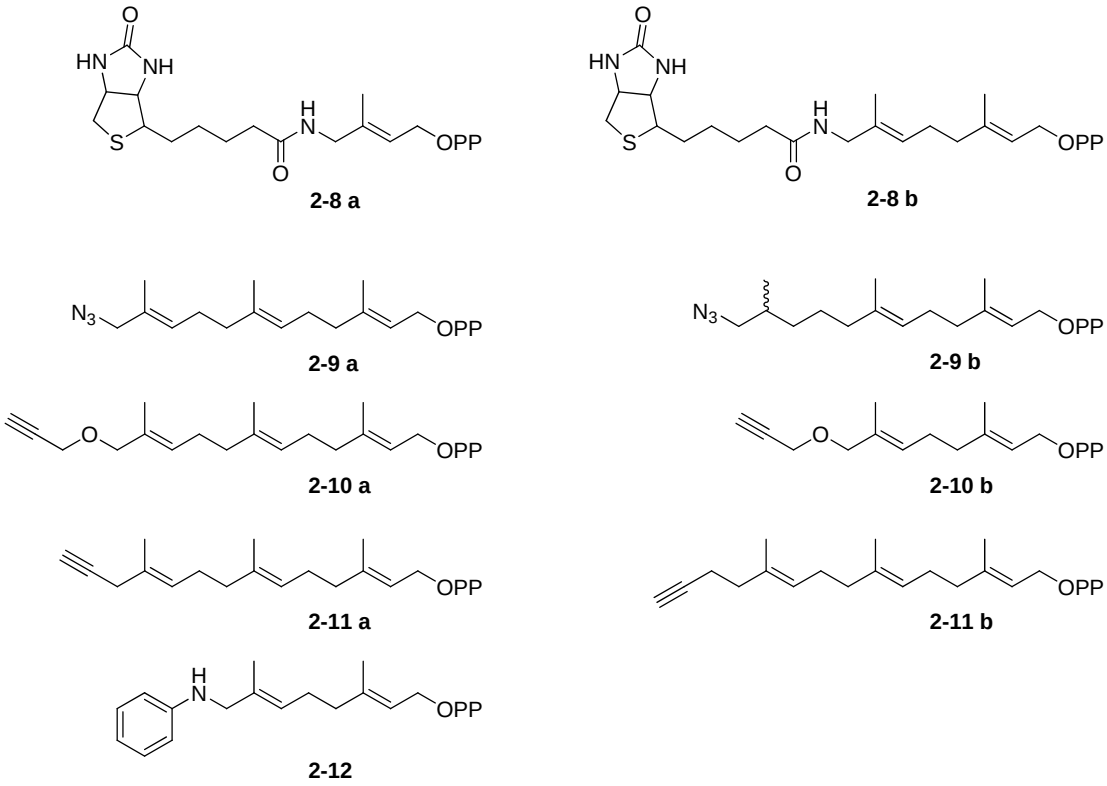


Figure 2.4 Probes for studying prenylome.

In order to detect prenylated proteins in cells, there must be a handle on the isoprenyl chain which allows researchers to observe those proteins or to capture them. Biotinylated derivatives (B-GPP, **2-8 b**) allowed the prenylated proteins to be enriched by streptavidin beads. The enriched prenylome can subsequently be analyzed by Western blot or MS.[97] The affinity of compound **2-8 b** for the PFTase, GGTase-I and GGTase-II is reduced compared to the native substrates but is comparable to that of other derivatized isopreniod analogues. Unfortunately, compound **2-8 b** is only a substrate for GGTase-II. No or only residual transfer activity could be detected for

PFTase and GGTase-I, respectively. Mutants of the PFTase (W102T_Y154T or W102T_Y154T_Y205T) and GGTase-I (F52Y_F53Y_Y126T) with expanded active sites were found to use compound **2-8 b** as a substrate. There are two disadvantages of using compound **2-8 b** as a probe to detect prenylome. First, if PFTase-prenylome or GGTase-I-prenylome is the goal, cells must be transfected to express mutant enzymes. This adds an extra complication. Second, it is known from crystallography that in the active site a peptide substrate and an isoprenoid substrate physically contact each other in some part[98] so the structure of the isoprenoid can influence the peptide specificity. The identified prenylome may not be the native prenylome if the isopreniod analogues are structurally very different from FPP or GGPP. Although compound **2-8 b** was demonstrated to be delivered to their cognate substrates faithfully and orthogonally by the engineered prenyltransferase, structural perturbation is always a concern. Biotinylated derivatives with shortened chain length (B-DMAPP, **2-8 a**) was again transferred only by GGTase-II and was transferred less efficiently than compound **2-8 b**.[93] This suggested that shortening of the isoprenoid chain resulted in reduction of the affinity for the enzyme.

A solution for the above mentioned problems is to use an isoprenoid analogue with small functional handle. After prenylation, a capture/labeling reagent was attached via a bioorthogonal chemical ligation technique. The choice of the ligation

reaction depends on the downstream application with the common ones being the 'click' reaction and Staudinger reaction. The advantage of this strategy is that different types of reagents can be ligated with one kind of isoprenoid analogue. A number of such reagents includes affinity labels (e.g., biotin, FLAG, etc.), reporter dyes and oligonucleotide tags. Isoprenoids of this kind include compound **2-9 a**, **2-9 b**, **2-10 a, 2-10 b, 2-11a** and **2-11 b**. They are all substrates of PFTase.

Growth of COS cells in the presence of compound **2-9 a** or its alcohol version resulted in incorporation of these azide-containing analogues into proteins. Chemoselective derivatization of prenylated proteins by a Staudinger reaction was done by using a biotinylated phosphine capture reagent.[99] The resulting protein conjugates can be specifically detected or affinity purified by streptavidin-linked horseradish peroxidase or agarose beads. Subsequent MS analysis identified a number of farnesylated proteins including those with potentially novel farnesylation motifs. It was reported that the azide of compound **2-9 a** can isomerize between C-12 and C-10.[100] Use of compound **2-9 b** should overcome this problem.

The rapid rate of the Cu(I)-catalyzed click reaction has made it the method of choice for many proteomic profiling protocols other than Staudinger reaction. However, background labeling occurs in the click reaction when the alkyne reagent is present in excess.[101,102] Significantly lower levels of non-specific reaction occur when

the azide partner is employed in high concentration. In fact, Distefano's group compared the labeling of prenylated proteins with TAMRA fluorophore in HeLa cells when incubating with compound **2-10 a** or alcohol version of **2-9 a**. SDS-PAGE and then in-gel fluorescence analysis showed that using compound **2-10 a** reduced background labeling.[101] 2-D gel electrophoresis of labeled proteins from HeLa cells grown in the presence of alcohol version of **2-10 a** or **2-10 b** followed by MS analysis allowed several labeled proteins to be identified.

Hang and coworkers have compared the labeling signals and patterns of **2-9 a, 2-10 a, 2-11 a** and **2-12 a**.[102] Jurkat cells were treated with alcohol version of **2-9 a, 2-10 a, 2-11 a** and **2-12 a**. Cell lysates were conjugated via CuAAC to azido-rhodamine or alkynyl-rhodamine. Lysates were separsated by SDS-PAGE and scanned for fluorescence. Robust signal over background was obtained using **2-10 a, 2-11 a** and **2-12 a** compared to **2-9 a**. This is consistent with other reports.[101] Labeling was stronger with **2-10 a** and **2-11 b** than **2-9 a** and **2-11 a**. In his research, **2-10** a was chosen for subsequent experiments because **2-10 a** was claimed to be readily synthesized and was stable compared to other alkyne-derivatized probes.[102,103]

Antibodies are useful both in the routine detection and immunoprecipitation of proteins with other post-translational modifications. However, it was reported that antibodies produced to detect farnesylated proteins showed cross-reaction with

geranylgeranylated proteins or others.[104-106] An alternative approach to detect prenylome is to produce antibodies against proteins modified with unnatural farnesyl analogues. In this way, it is less likely that the raised antibodies will cross-react with other proteins. Since compound **2-12** (AGPP) is also an effective alternative substrate for PFTase, Spielman and coworkers have produced antibodies specific for the anilinogeranyl moiety.[107] The alcohol version of compound **2-12** was incorporated into cellular proteins in a PFTase dependent manner. In a subsequent report, 2-D electrophoresis and Western blot with these antibodies revealed farnesylated proteome in leukemia cells.[108]

Chapter 3: Solid-Phase Synthesis of C-terminal Peptide Libraries

3.1 Introduction

Early investigations of Ca_1a_2X substrate preferences of prenyltransferases involved mutations of the C-terminal sequences of Ras proteins. Individual purified proteins were assayed using [^{3}H]-FPP or [^{3}H]GGPP as the isoprenoid substrate.[109] This method is inconvenient and labor-intensive. Subsequently, a fluorescence assay was developed to study peptide specificity of PFTase and GGTase-I[110,111]. That involves the preparation of dansylated Ca_1a_2X-box containing peptides, and measurement of dansyl-group fluorescence that increases upon prenylation.[112] However, this method still requires the synthesis, purification and analysis of individual peptides which is a time-consuming process. Combinatorial approaches that use one-bead-one-compound libraries or peptide arrays provide possible strategies for increasing the number of sequences that can be examined. However, studies of protein prenylation present a number of challenges to implementing such methods. First, prenylation is a post-translational modification that occurs at the C-terminus of a protein. Since solid phase peptide synthesis is typically performed from the C- to N-terminus, more complex synthetic procedures are required to prepare libraries with free C-termini. Next,

prenyltransferases are large multimeric enzymes (~80 kDa or larger) whose size reduces their ability to penetrate into the resins typically employed for solid phase synthesis; the enzymes responsible for the subsequent processing of prenylated proteins are membrane-bound presenting additional challenges. Finally, isoprenoids are not intrinsically chromogenic and are sensitive to acidic conditions and electrophiles; those features make the enzymatic addition to peptides on resin difficult to monitor and the subsequent cleavage and sequence analysis more complex. In this chapter, we describe the synthesis of a library of peptides containing free C-termini on cellulose membranes. The resulting library was then used to probe the specificity of protein farnesyltransferase from *S. cerevisiae* (yPFTase).

3.2 Synthesis of a peptide library with free C-termini and its screening against yPFTase

Our approach for the preparation of peptide libraries with free C-termini (Figure 3.1) is based on the pioneering work of Kania, Zuckermann and Marlowe[31] and key improvements contributed by others.[32,46] A four residue linker BBLL was initially added on to amine-functionalized solid surfaces. The resulting peptide was then coupled to the γ-carboxyl group of Fmoc-Glu-ODmab; this

resulted in the installation of an orthogonally (Dmab) protected carboxylate that could be used for subsequent on-resin cyclization. Fmoc removal and coupling of an HMPA linker followed by ester bond formation produced a depsipeptide containing an acid-labile ester linkage. Seven residues were then sequentially added ending with a linker to permit cleavage of the final product from the resin at the end of the synthesis. After Dmab removal from the internal Glu, the resulting free amino- and carboxyl-groups were coupled to yield a cyclic peptide on resin; treatment with modified Reagent K caused cleavage of the internal ester linkage and afforded the desired linear peptide with a free C-terminus.

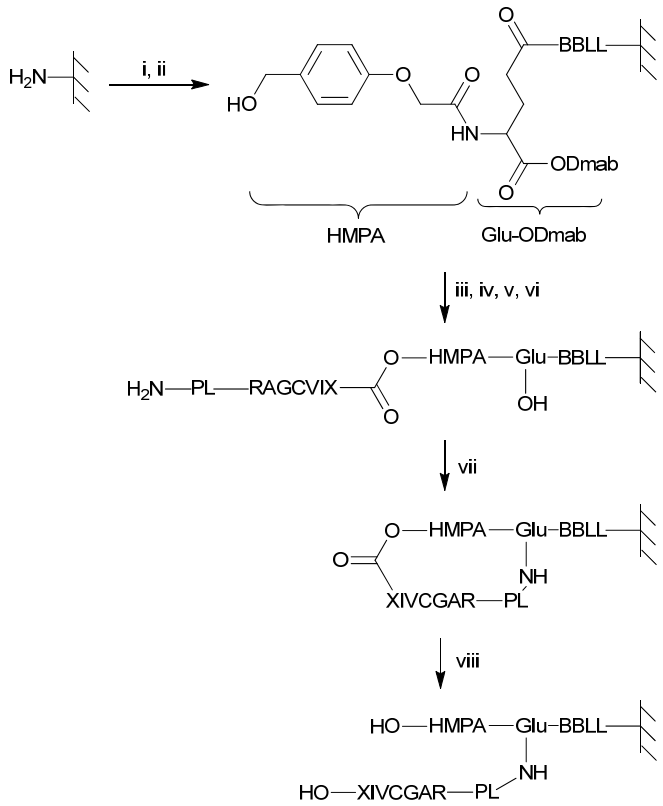


Figure 3.1 Synthesis of C-terminal peptides. Reagents and conditions: (i) standard DIC coupling of Fmoc-Aa (2×), then capping, then 20% piperidine, then BPB staining; (ii) standard DIC coupling of HMPA (2×); (iii) 0.5 M Fmoc-Aa and 0.5 M 6-Cl-HOBt in DMF/CH_2Cl_2, 0.5 M DIC in CH_2Cl_2, 0.2 M DMAP in CH_2Cl_2 (6-8×), then capping, then 20% piperidine, then BPB staining; (iv) standard DIC coupling of Fmoc-Aa (2×), then capping, then 20% piperidine, then BPB staining; (v) 0.5 M photolysis linker, 0.5 M Et_3N in DMF (3-4×); (vi) 2% hydrazine; (vii) 0.05 M BOP, 0.05 M 6-Cl-HOBt and 0.1 M DIEA in DMF (2×); (viii) modified reagent K. PL represents a cleavable linker so that peptides can be cleaved from resin and analyzed by MS. B represents β-alanine.

Initially, we planed to use Met as the linker in the PL position (Figure 3.1) and cleave the products after enzymatic prenylation from the resin using CNBr for subsequent analysis; previous reports suggested that prenylated cysteine was stable to CNBr. However, we found that after chemical farnesylation of the resin-bound peptide, treatment with CNBr resulted in significant amounts of prenyl group

cleavage. Additionally, CNBr cleavage suffers from slow kinetics and results in the formation of side products observed via MS (Figure 3.2).

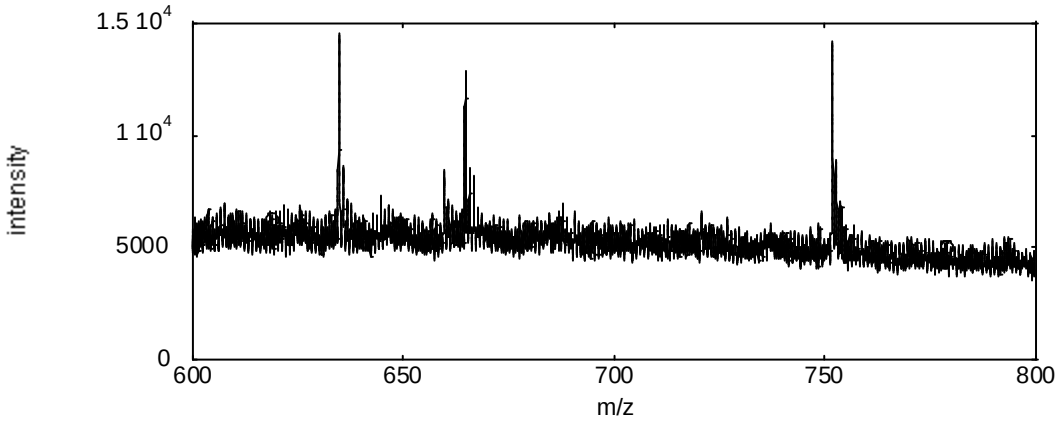


Figure 3.2 MALDI analysis of peptide prepared using Met as the linker. The resin was treated with 100 mg/mL CNBr in 70% TFA overnight followed by solvent removal *in vacuo*. The peptide (RGCVLS) was re-dissolved in CH_3CN/H_2O (v/v = 5/95) with 0.1% TFA for subsequent MS analysis.

To circumvent this problem, we initially examined the use of the photocleavable linker shown in Figure 3.3a. Incorporation of that structure into a peptide prepared via the route outlined in Figure 3.1 followed by chemical farnesylation and photocleavage produced the desired free peptide. However, analysis of that material was complicated due to the fragmentation of the nitroso ketone moiety that remained linked to the final peptide[113] (Figure 3.4); while not problematic in the case of a single peptide, it could significantly complicate

matters in the context of library analysis.

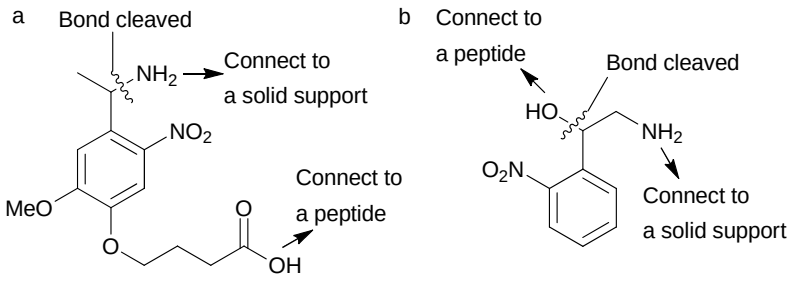


Figure 3.3 Photocleavable linkers used for peptide synthesis in this study.

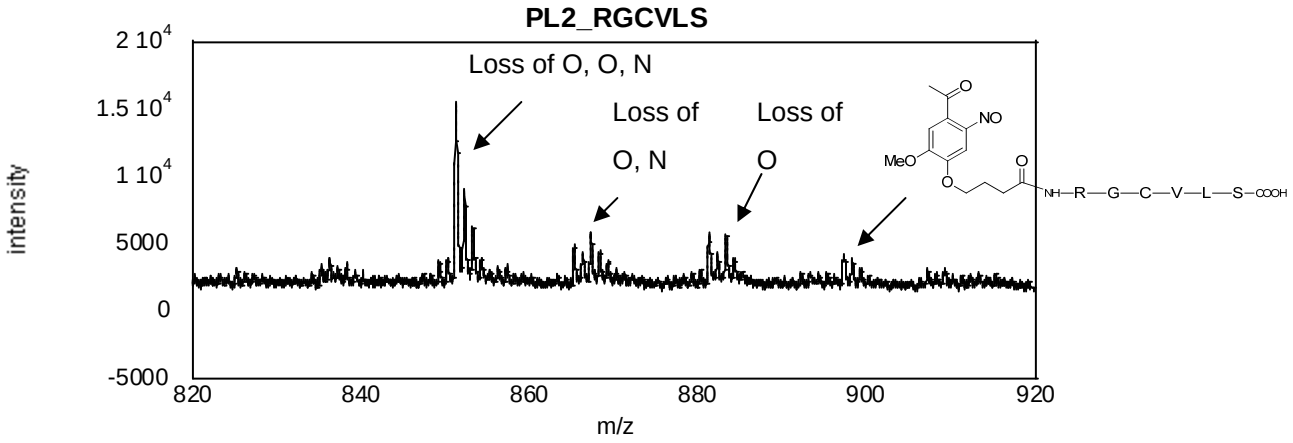


Figure 3.4 MALDI analysis of peptide prepared using first generation photocleavable linker (structure shown in Figure 3.3a). The resin was treated with UV light (365 nm) in CH_3CN/H_2O (v/v = 5/95) followed by washing with CH_3CN and MeOH. The combined solvents were removed *in vacuo* and the resulting peptides (PL-RGCVLS) were re-dissolved in CH_3CN/H_2O (v/v = 5/95) with 0.1% TFA for subsequent MS analysis.

Accordingly, a new photocleavablelinker, shown in Figure 3.3b, was designed and prepared (see Scheme S3.1). The resulting residue was incorporated into

peptides as “PL” using the strategy outlined in Figure 3.1. After chemical farnesylation, photolysis of the resin resulted in clean formation of the desired peptide lacking the nitroso ketone substructure (which remained resin bound) (Fugure 3.5).

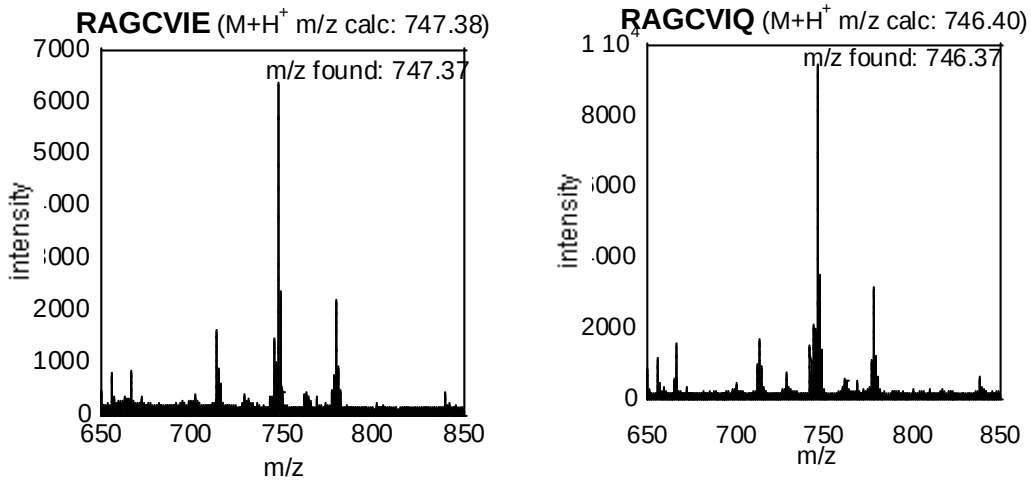
With a workable strategy for the synthesis and analysis of relevant C-terminal peptides in hand, we next began to examine the enzymatic prenylation of these molecules. Unfortunately, treatment of TentaGel beads terminating in the sequence CVIA with PFTase and FPP for prolonged reaction time did not result in the formation of any detectable farnesylated product by MS after photocleavage; experiments employing a more hydrophilic PEGA resin as the solid support for synthesis gave similar negative results. Interestingly, after using controlled pore glass (CPG) with a large pore size (300 Å) as the solid support for synthesis, it was possible to detect the enzymatic product. Those results suggest that the large size of PFTase, a heterodimer of 80-90 kDa, limits the ability of the enzyme to penetrate into the interior of many conventional supports used for SPPS.[114] While the use of CPG for peptide synthesis appeared to provide a viable solution to this problem, we were concerned that the much lower peptide loading obtainable with this material would be problematic when trying to identify specific peptide sequences present on a single bead in a one-bead-one-compound library. Thus we elected to pursue a parallel

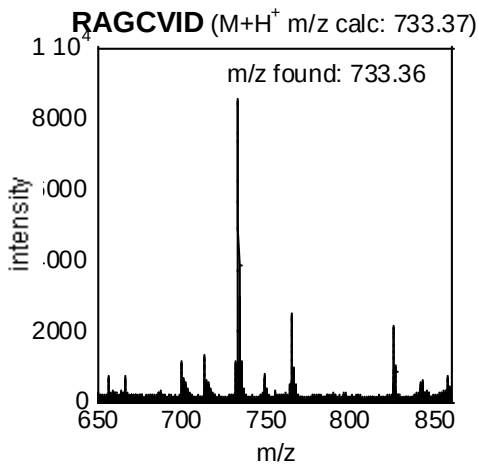
synthesis strategy where peptides were synthesized on amine-functionalized cellulose filters (SPOT synthesis). Gratifyingly, synthesis of C-terminal peptides ending in CVIA using the synthetic procedure described above proceeded smoothly and analysis after enzymatic prenylation with FPP showed essentially quantitative conversion of the peptide to the corresponding farnesylated product.

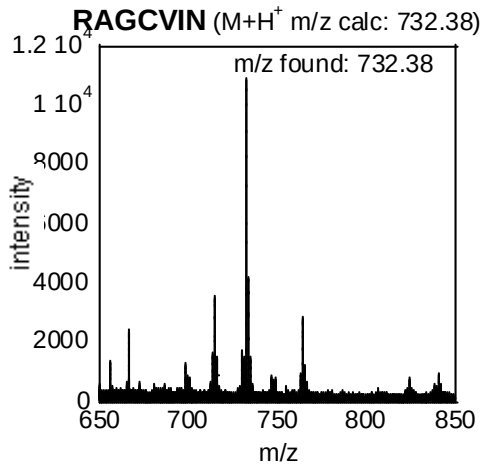
While MS analysis of peptides released from the resin after photocleavage worked well for the detection of prenylated peptides in the development of this methodology, such an approach is not ammenable to the screening of large libraries of peptides. Originally, we planned to use a fluorescent FPP analogue to accomplish this[90-92]; in such a strategy, spots containing actual peptide substrates should be fluorescently labeled in the course of enzymatic prenylation and be easily distinguished from those that are not substrates. However, low fluorescence of the membranes made the location of fluorescent spots difficult to distinguish. Thus, an alternative approach was used in which an alkyne-containing analogue of FPP[101,115] was initially transferred followed by click reaction with biotin-azide and subsequent detection with streptavidin-AP using a chromogenic substrate. Hence, that procedure was used for all library screening described here. It should be noted that the alkyne-containing analogue employed here has a V_{max} value comparable to that of FPP[115] and that screening was performed under saturating substrate

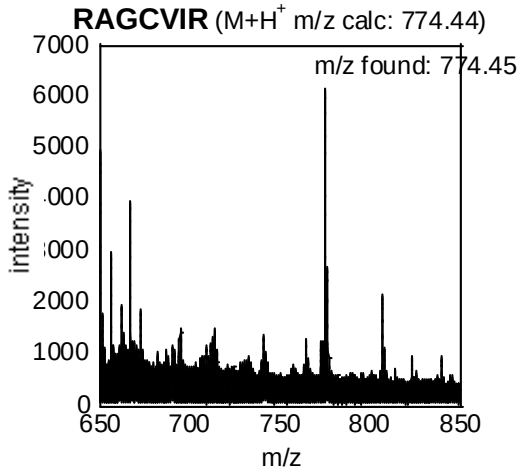
conditions.

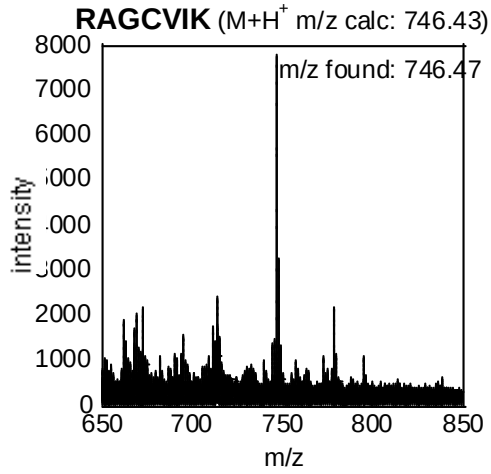
Having validated the method for a single peptide, we prepared a 20-member library via manual SPOT synthesis[2] in which the C-terminal residue of the CVIX sequence was varied to contain any of the 20 proteogenic amino acids. In the first experiment, upon completion of the synthesis, individual spots were excised from the membranes and cut in half. In each case, one half-spot was subjected to enzymatic prenylation with OPP-Far-alkyne followed by click reaction with biotin-azide and visualization with SP-AP. The remaining half-spots were used to confirm that the desired peptides were present on the cellulose filters (Figure 3.5).

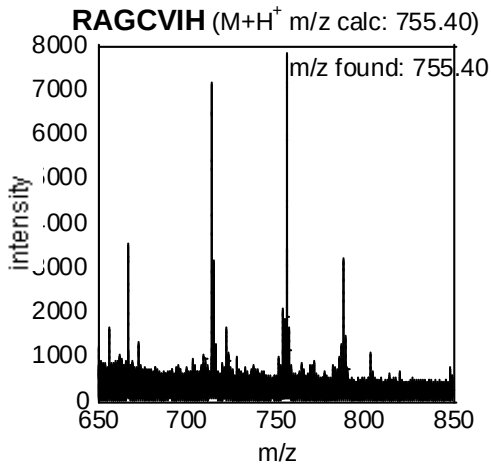


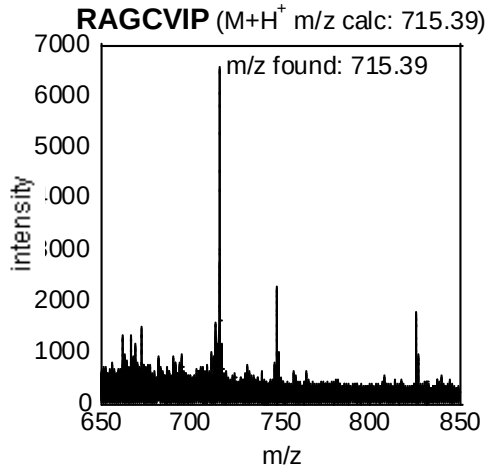

RAGCVID (M+H+ m/z calc: 733.37)
m/z found: 733.36
intensity
m/z

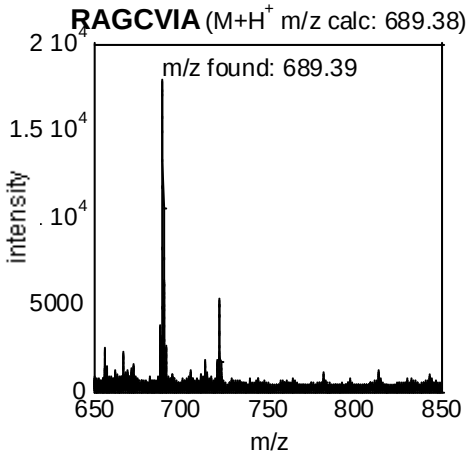

RAGCVIN (M+H+ m/z calc: 732.38)
m/z found: 732.38
intensity
m/z

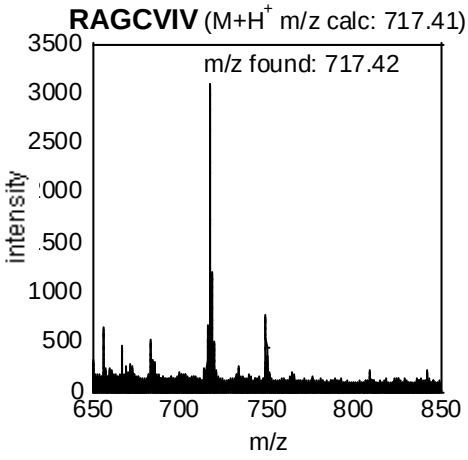

RAGCVIR (M+H+ m/z calc: 774.44)
m/z found: 774.45
intensity
m/z

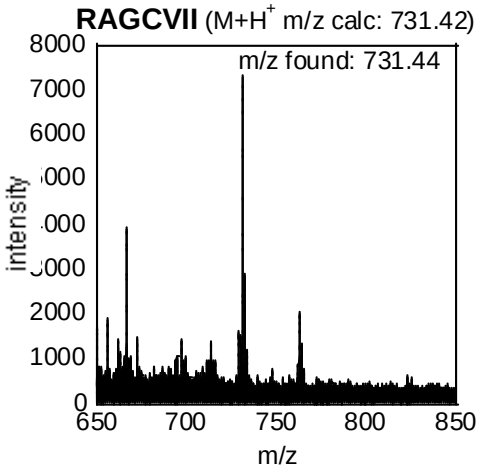

RAGCVIK (M+H+ m/z calc: 746.43)
m/z found: 746.47
intensity
m/z

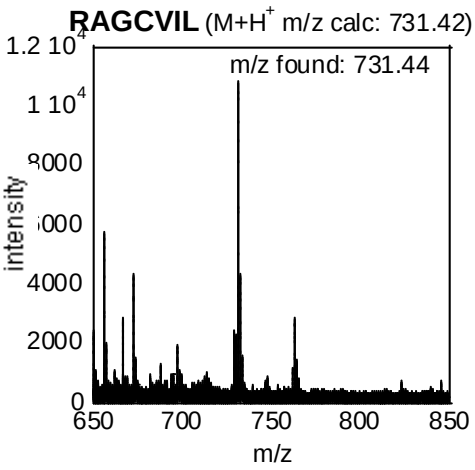

RAGCVIH (M+H+ m/z calc: 755.40)
m/z found: 755.40
intensity
m/z

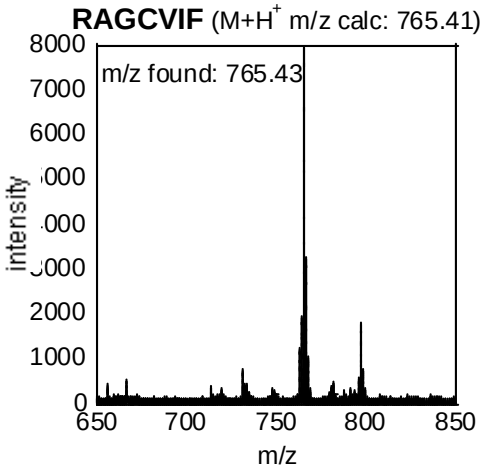

RAGCVIP (M+H+ m/z calc: 715.39)
m/z found: 715.39
intensity
m/z

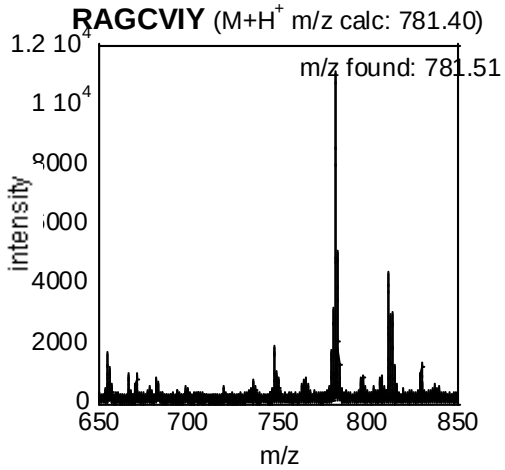

RAGCVIA (M+H+ m/z calc: 689.38)
m/z found: 689.39
intensity
m/z


RAGCVIV (M+H+ m/z calc: 717.41)
m/z found: 717.42
intensity
m/z


RAGCVII (M+H+ m/z calc: 731.42)
m/z found: 731.44
intensity
m/z


RAGCVIL (M+H+ m/z calc: 731.42)
m/z found: 731.44
intensity
m/z


RAGCVIF (M+H+ m/z calc: 765.41)
m/z found: 765.43
intensity
m/z


RAGCVIY (M+H+ m/z calc: 781.40)
m/z found: 781.51
intensity
m/z

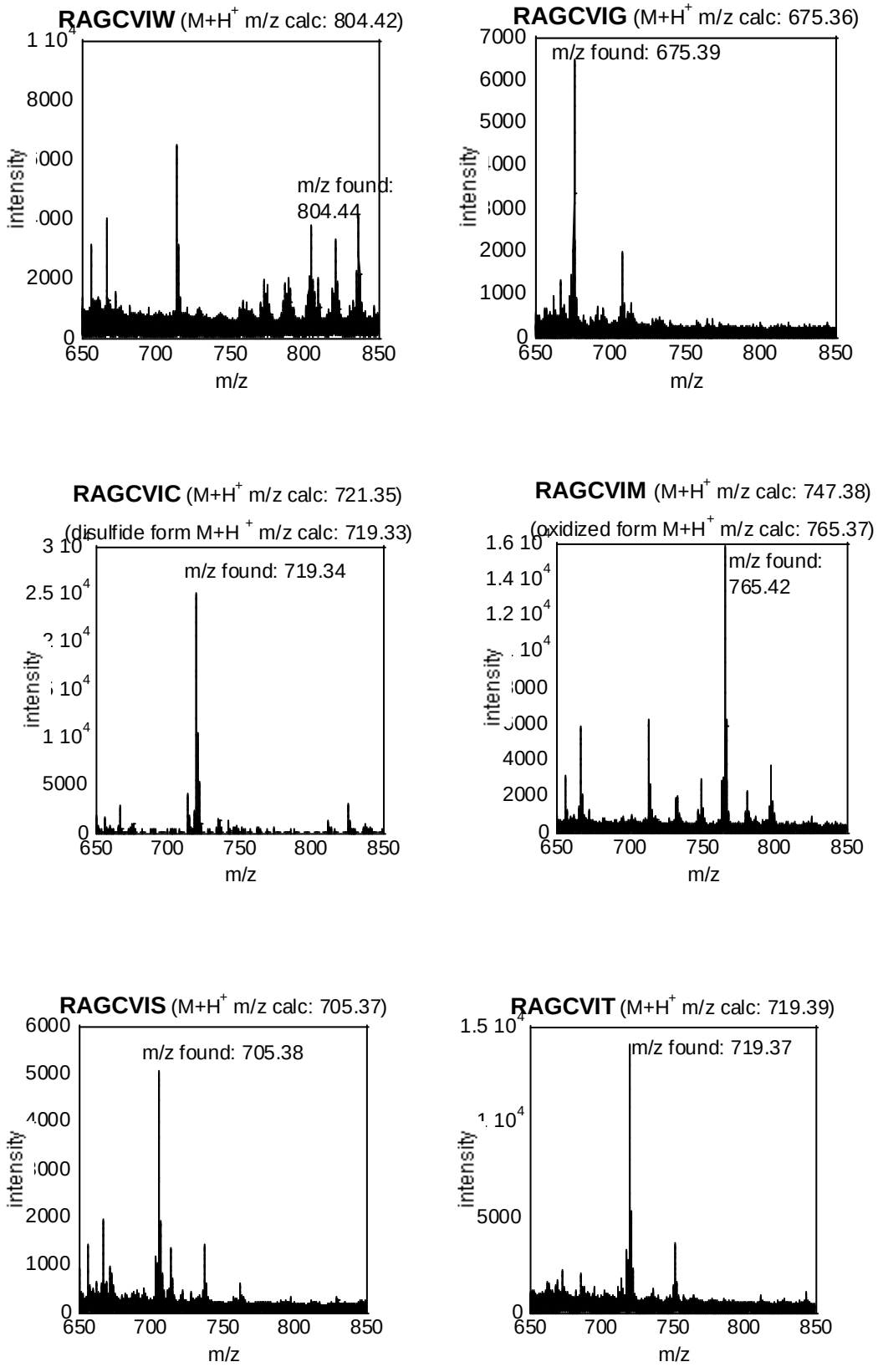
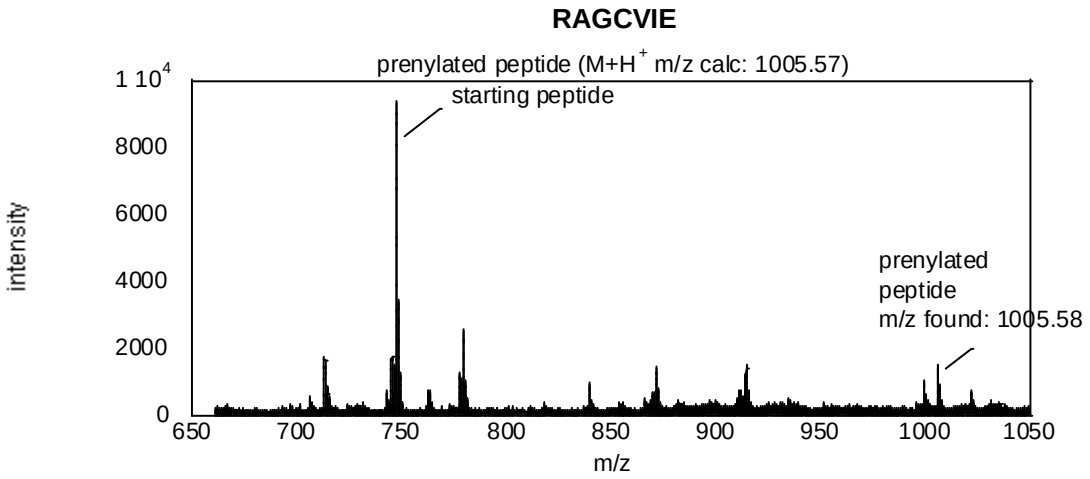


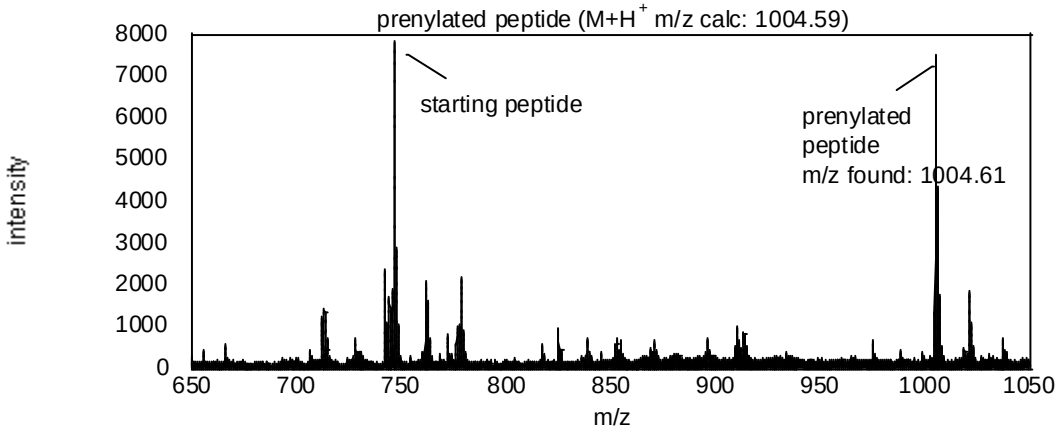
Figure 3.5 MALDI analysis of peptides produced by SPOT synthesis performed to verify the

production of the desired product. The photocleavable moiety shown in Fig. 3B (main text) was used as the linker. The membranes were treated with UV light (365 nm) in CH_3CN/H_2O (v/v = 5/95). The resins were washed with CH_3CN and MeOH. The combined solvents were removed *in vacuo*. The peptides (RAGCVIX) were re-dissolved in CH_3CN/H_2O (v/v = 5/95) with 0.1% TFA.

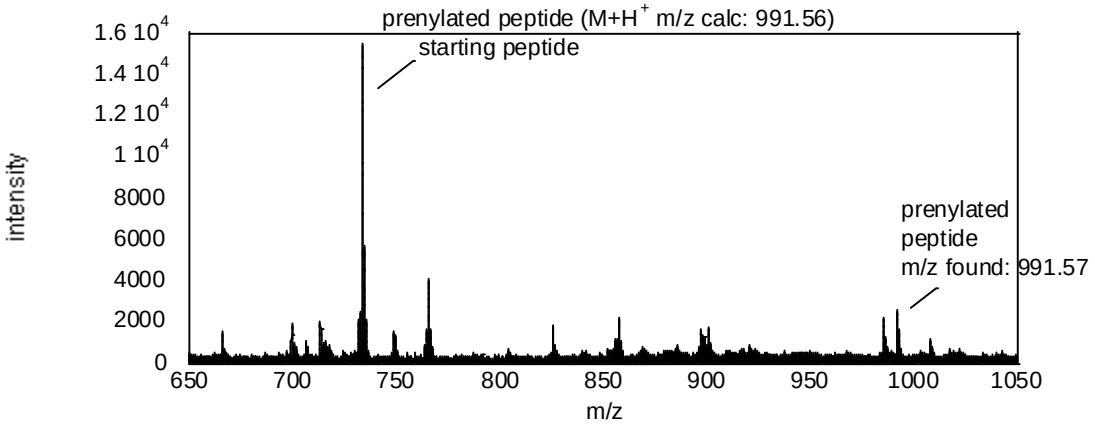
In a secondary experiment for evaluating whether the peptides were prenylated, individual spots were cut in half after completion of the enzymatic prenylation. One half-spot was subjected to colorization as described above. Peptides from the other half-spots were photocleaved followed by analysis via MALDI-MS. (Figure 3.6).



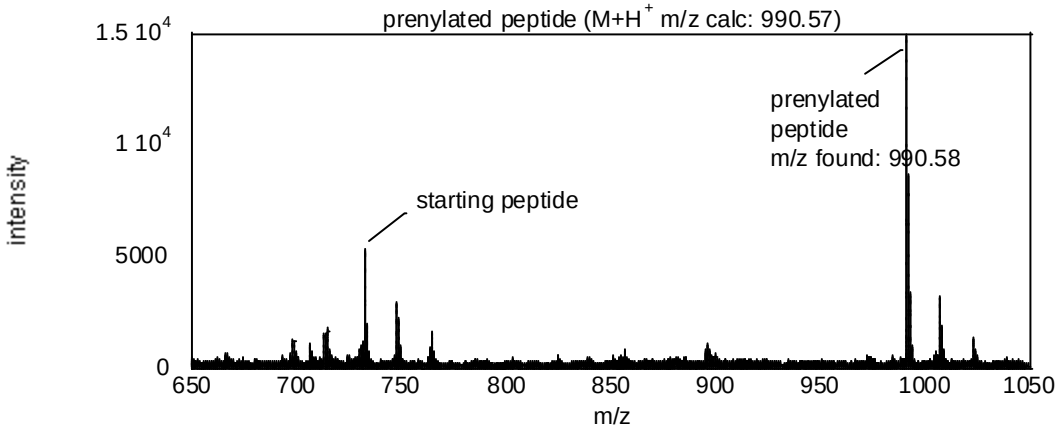
RAGCVIQ

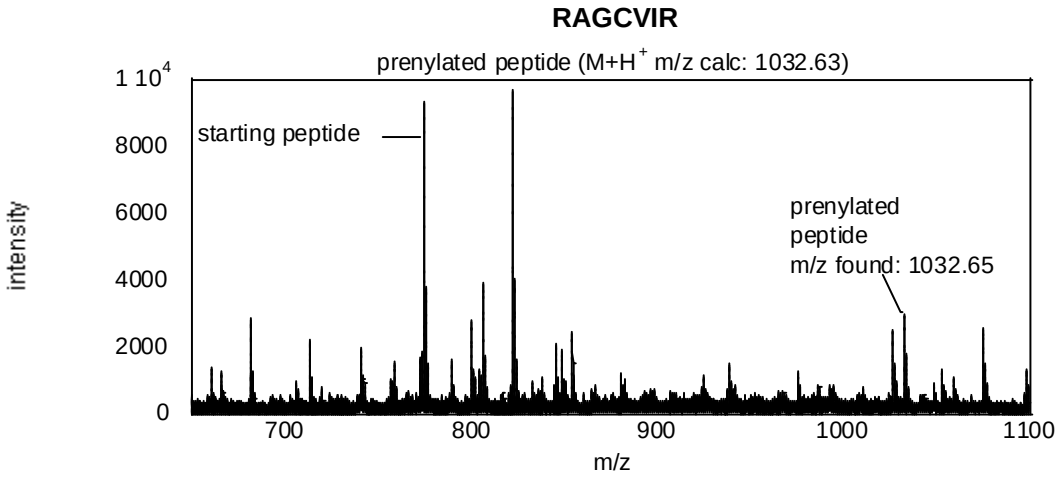


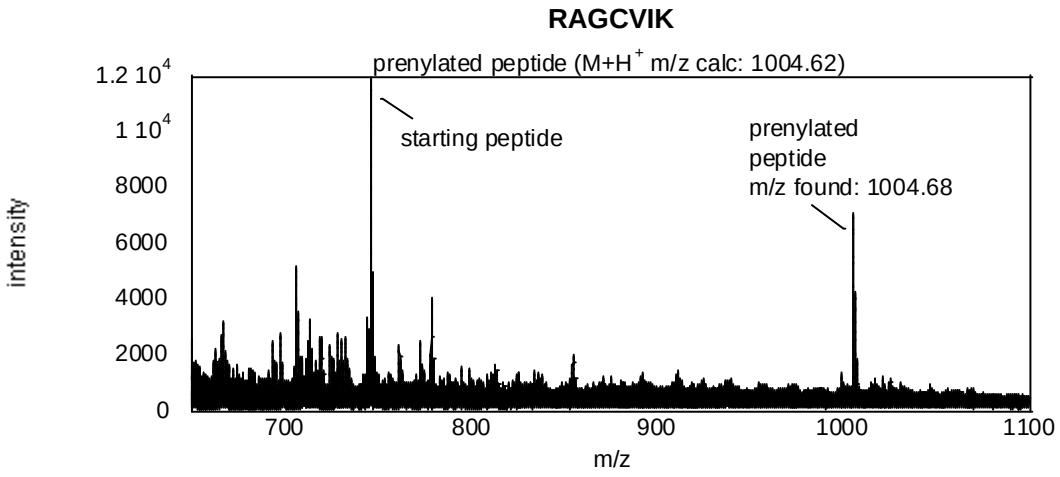
RAGCVID

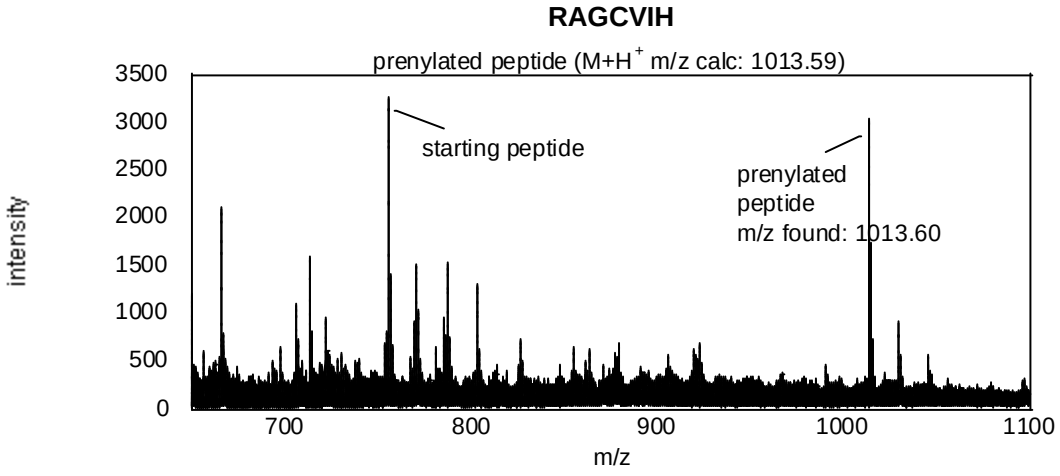


RAGCVIN

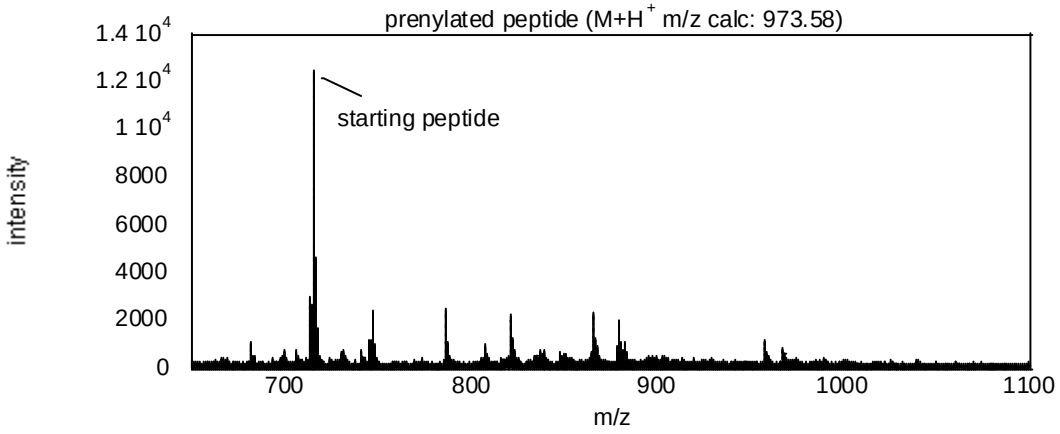



RAGCVIR
prenylated peptide (M+H+ m/z calc: 1032.63)
1 10^4
8000
6000
4000
2000
0
intensity
starting peptide
prenylated
peptide
m/z found: 1032.65
700
800
900
1000
1100
m/z

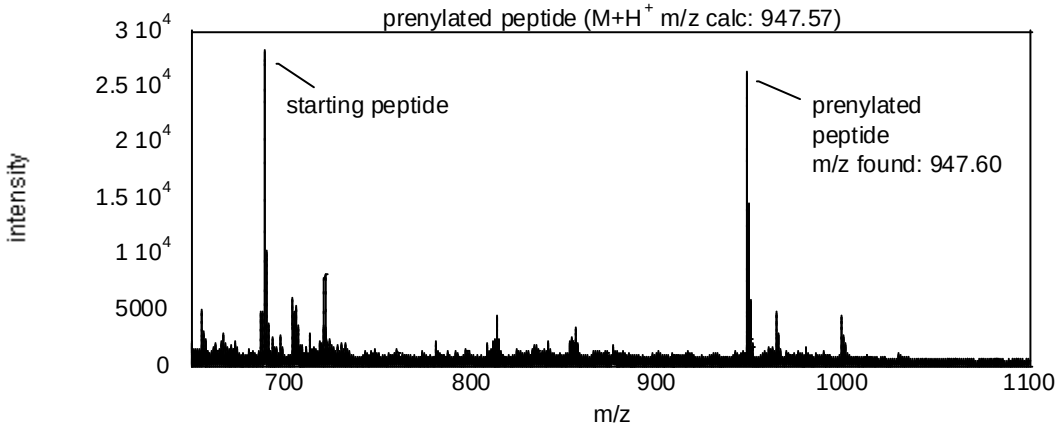

RAGCVIK
prenylated peptide (M+H+ m/z calc: 1004.62)
1.2 10^4
1 10^4
8000
6000
4000
2000
0
intensity
starting peptide
prenylated
peptide
m/z found: 1004.68
700
800
900
1000
1100
m/z


RAGCVIH
prenylated peptide (M+H+ m/z calc: 1013.59)
3500
3000
2500
2000
1500
1000
500
0
intensity
starting peptide
prenylated
peptide
m/z found: 1013.60
700
800
900
1000
1100
m/z

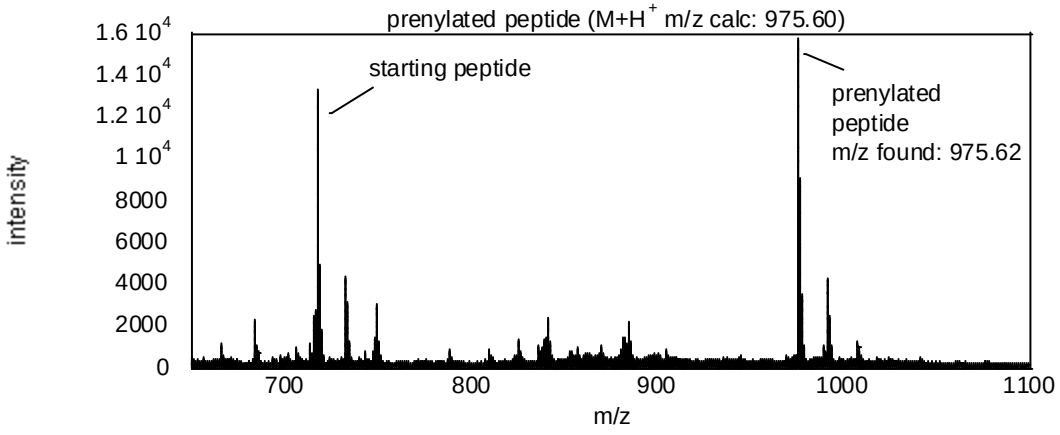
RAGCVIP

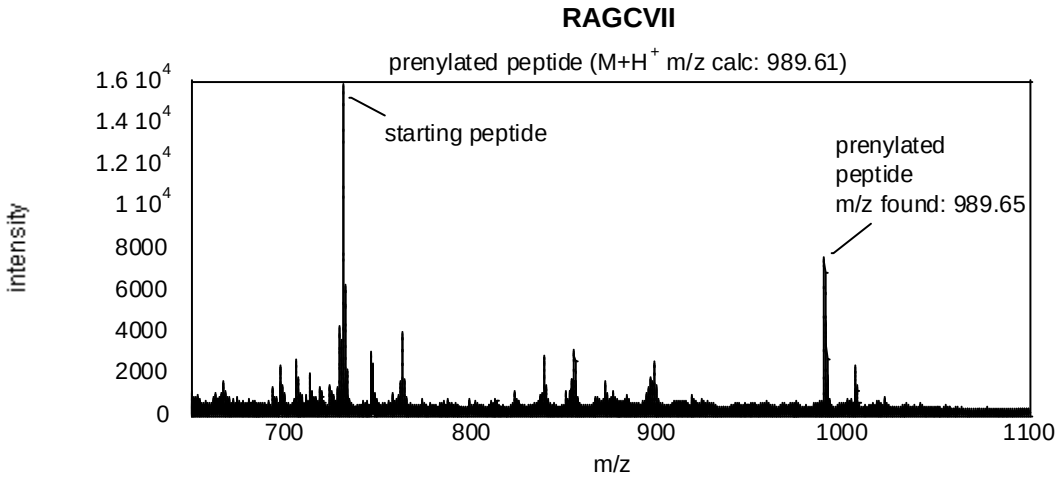


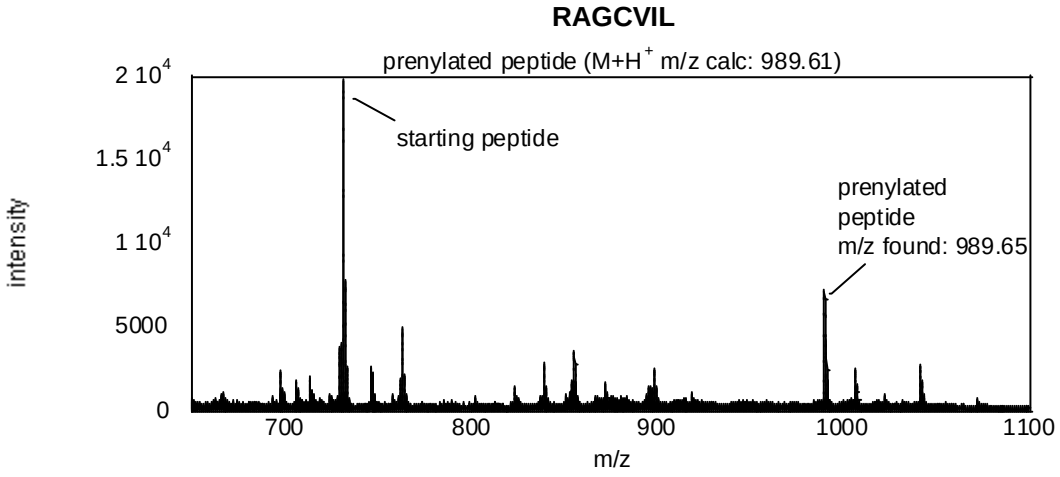
RAGCVIA

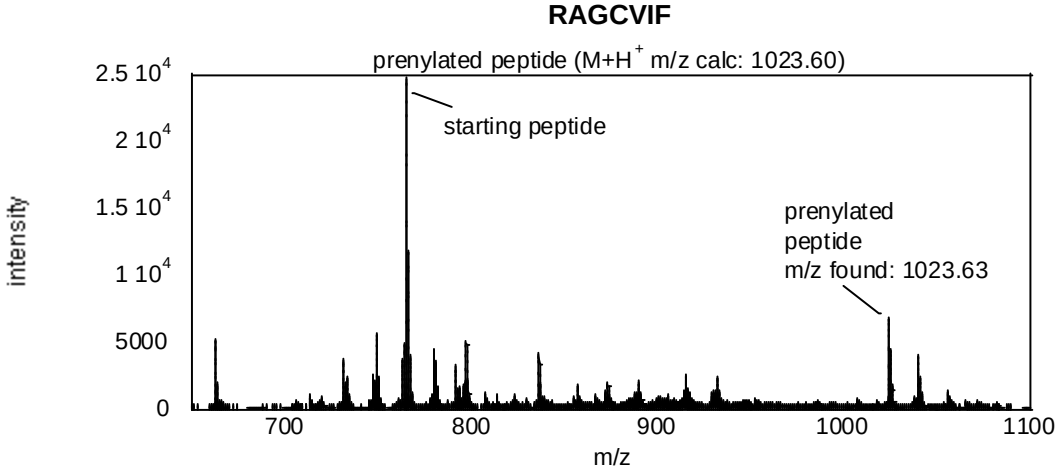


RAGCVIV

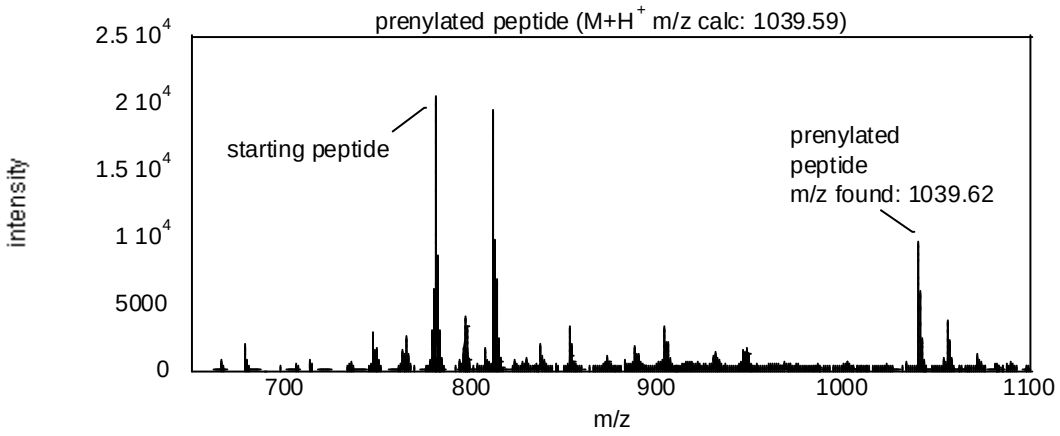



RAGCVII
prenylated peptide (M+H+ m/z calc: 989.61)
starting peptide
prenylated peptide m/z found: 989.65
intensity
m/z
1.6 10^4
1.4 10^4
1.2 10^4
1 10^4
8000
6000
4000
2000
0
700
800
900
1000
1100

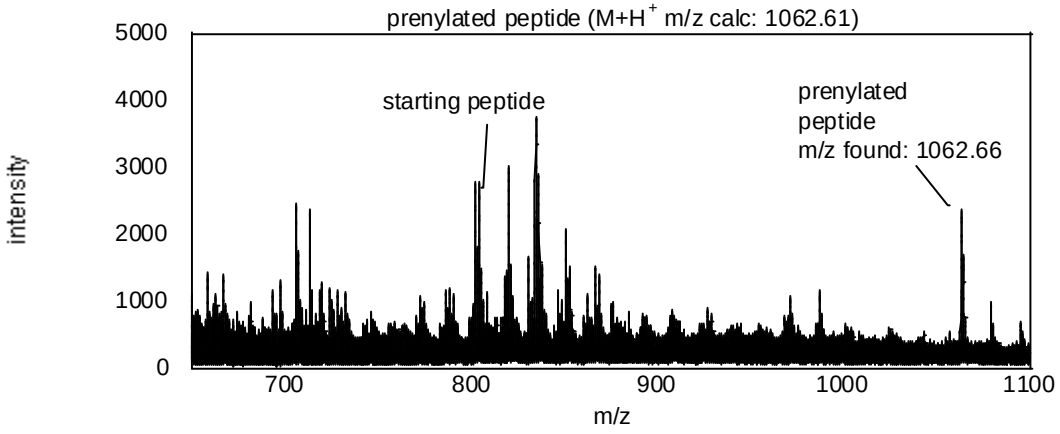

RAGCVIL
prenylated peptide (M+H+ m/z calc: 989.61)
starting peptide
prenylated peptide m/z found: 989.65
intensity
m/z
2 10^4
1.5 10^4
1 10^4
5000
0
700
800
900
1000
1100


RAGCVIF
prenylated peptide (M+H+ m/z calc: 1023.60)
starting peptide
prenylated peptide m/z found: 1023.63
intensity
m/z
2.5 10^4
2 10^4
1.5 10^4
1 10^4
5000
0
700
800
900
1000
1100

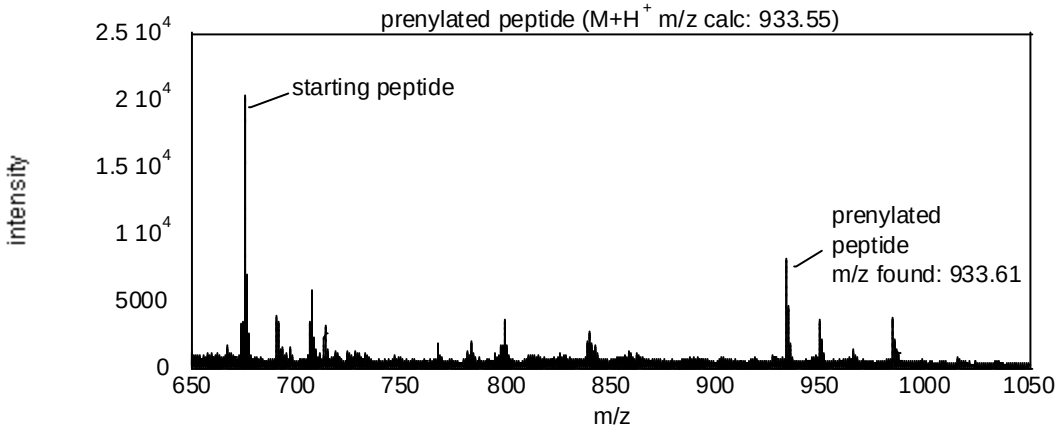
RAGCVIY



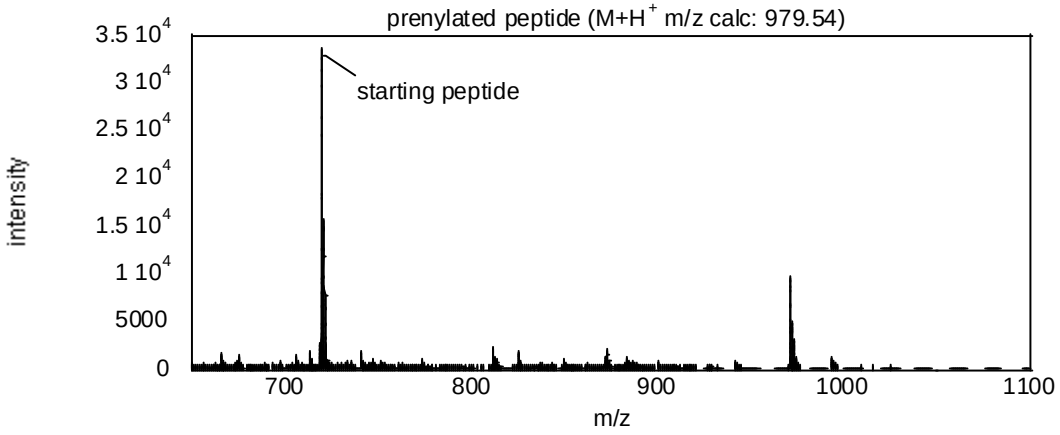
RAGCVIW



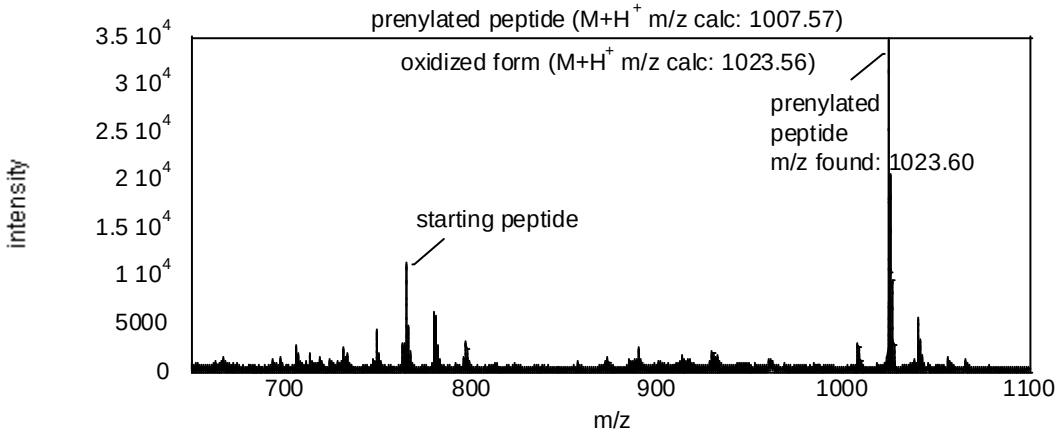
RAGCVIG



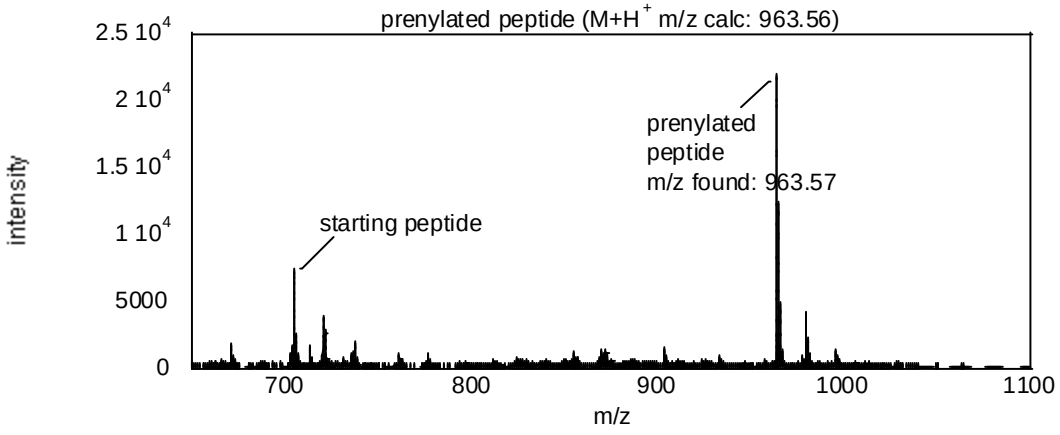
RAGCVIC



RAGCVIM



RAGCVIS



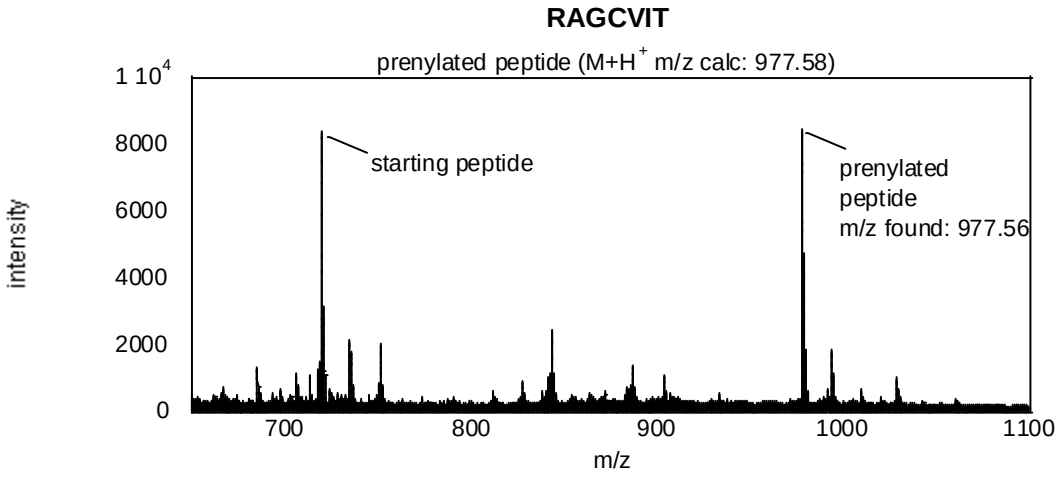


Figure 3.6 MALDI analysis of peptides produced by SPOT synthesis after enzymatic farnesylation performed to confirm successful enzymatic processing. The photocleavable moiety shown in Fig. 3B (main text) was used as the linker. After enzymatic reaction, the membranes were treated with UV light (365 nm) in CH_3CN/H_2O (v/v = 5/95) to release the peptides. The membranes were washed with CH_3CN and MeOH and the combined washes were evaporated *in vacuo*. The peptides (RAGCVIX) were re-dissolved in CH_3CN/H_2O (v/v = 5/95) with 0.1% TFA.

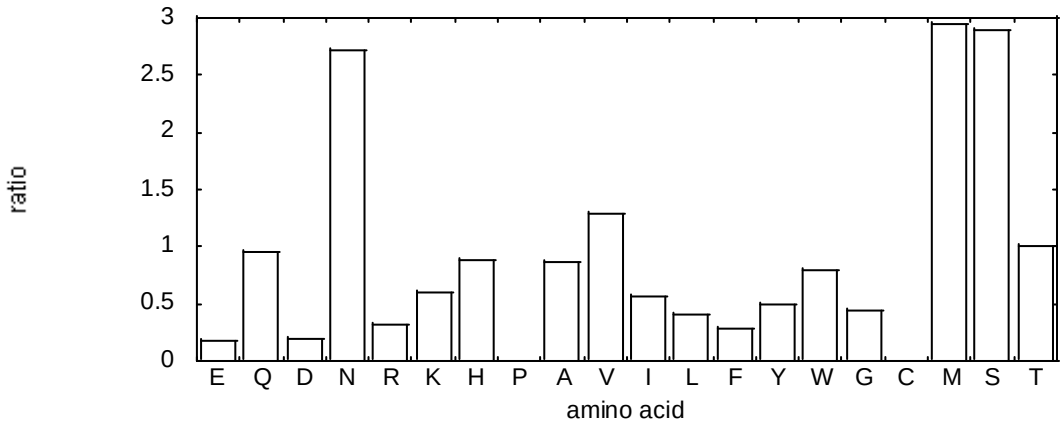


Figure 3.7 MS intensity ratio of prenylated peptides versus unprenylated peptides from Figure S4 performed to determine the extent of enzymatic prenylation of peptides prepared via SPOT synthesis.

For analysis of the spots subjected to prenylation and visualization with SP-AP, the intensity of the color was normalized to that obtained with CVIQ (Figure 3.8 and Figure 3.9). Based on that analysis, eight residues at the X position gave signals comparable to, or greater than A, including E, Q, D, N, G, M, S, and T. Of those, D and E did not show levels of prenylation (ratio of prenylated peptide to unprenylated peptide < 0.5, Figure 3.7) in the secondary MALDI-MS-based screen. Ds-GCVID also showed low reaction rate in a continuous spectrofluorimetric assay (Table 3.1). If streptavidin-horseradish peroxidase (SP-HRP) was used in lieu of SP-AP for screening, peptides containing D and E at the X position manifested much less signal suggesting that those sequences are not efficient substrates. However, the overall level of background staining obtained with SP-HRP was significantly greater than that observed with SP-AP limiting the dynamic range of the the screen. Hence, our preferred procedure is to employ SP-AP as the primary screen and then use SP-HRP in subsequent screening when necessary.

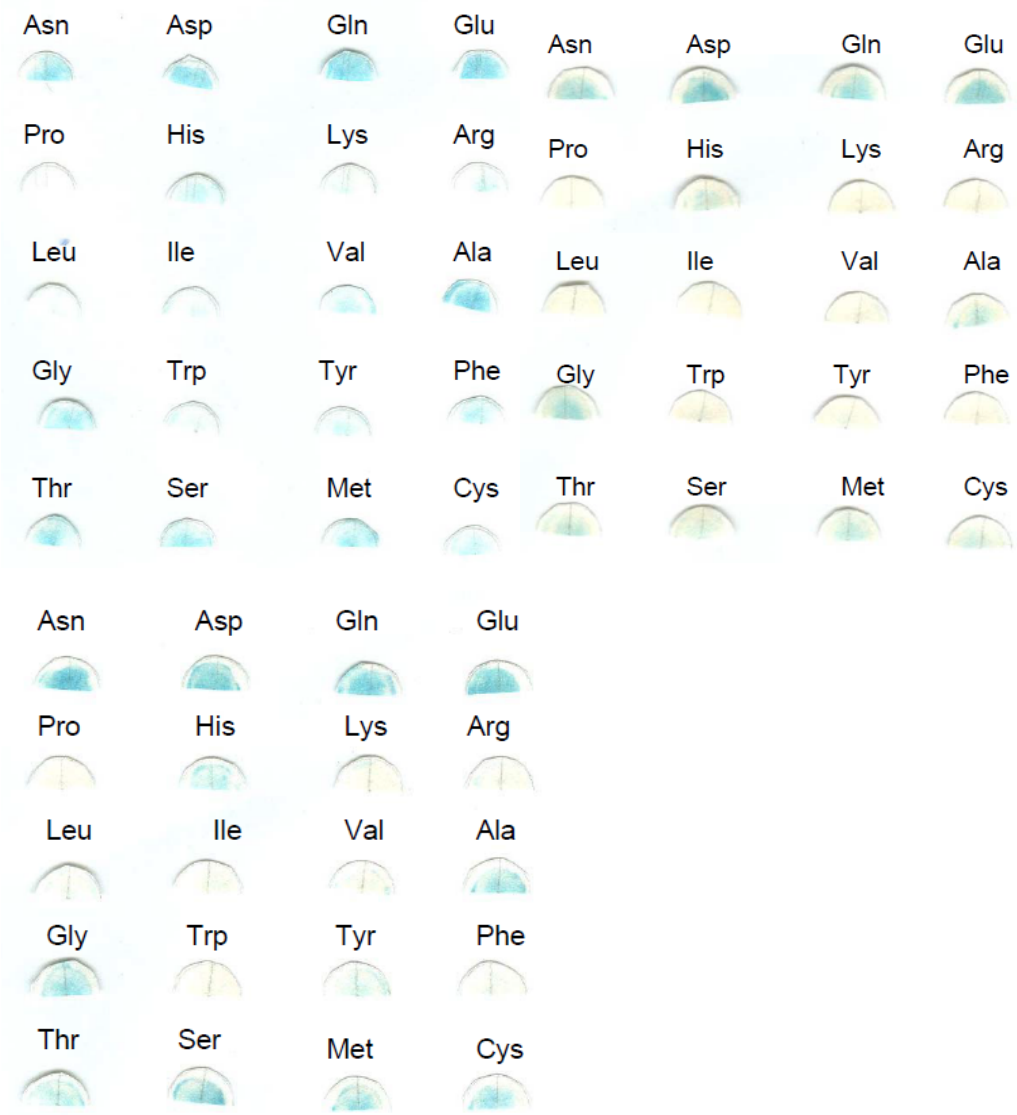


Figure 3.8 Screening of enzymatic prenylation of an RAGCVIX library. Each SPOT membrane was prenylated with OPP-Far-alkyne by PFTase, clicked with biotin-azide and then visualized by SP-AP in BCIP solution. Membranes from three separate library syntheses and screenings are shown.

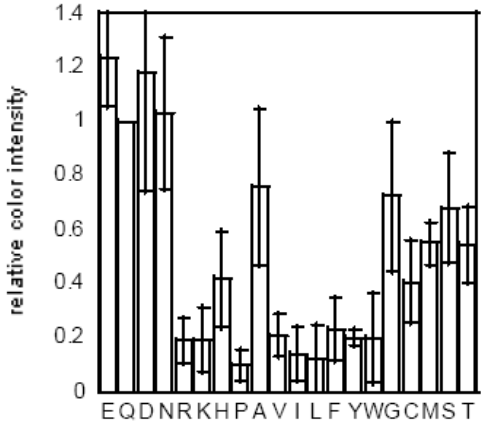


Figure 3.9 Evaluation of the extent of farnesylation of a RAGCVIX library of peptides. After synthesis, the peptides were subjected to enzymatic prenylation with an alkyne-containing analogue of FPP followed by click reaction with biotin-azide and subsequent reaction with streptavidin-AP using a chromogenic substrate. Color intensity was quantified by Image J software. The library was synthesized and screened three times and the average color intensities from the three trials are plotted here along with the SD depicted by the error bars. For comparison, the intensity was normalized relative to that observed with Gln at the X position.

v (uM/min)	Ds-GCVIN	DsGCVIA	Ds-GCVIT
	3.60 ± 0.25	2.39 ± 0.25	2.28 ± 0.10
	Ds-GCVID	Ds-GCVIK	Ds-GCVLL
	0.033 ± 0.003	0.048 ± 0.009	0.032 ± 0.004

Table 3.1 Evaluation of Ds-GCaaX peptides versus yFTase. Assays were performed using 2.4 μM dansylated peptide, 70 nM FTase, 10 μM OPP-Far-alkyne, in 200 mM Tris, pH 7.5, 5 mM DTT, 5 mM $MgCl_2$, 50 μM $ZnCl_2$, and 0.040 % (*w/v*) *n*-dodecyl-β-D-maltoside.

To extend these results, we next employed automated SOPT synthesis to build a 17x20 CVa_2X library with the X being one of the 20 proteogenic amino acids except Q, P, Y and a_2 being one of the 20 proteogenic amino acids. The procedure for peptide synthesis was similar except at step (iii) where 1,1'-carbonyldiimidazole in DMF was used for coupling.[32] Mass spectrometric

analysis of selected peptides from the library showed the presence of the desired peptides with the exception of peptides containing W at the X position where some side products were again observed (similar to the results from the manual synthesis noted above). Synthesis, enzymatic prenylation and screening were performed in duplicate using the method described above. The intensity of the color was normalized to that obtained with CVIN and color coded with the most intense signals shown in red (Figure 3.10 and Figure 3.11).

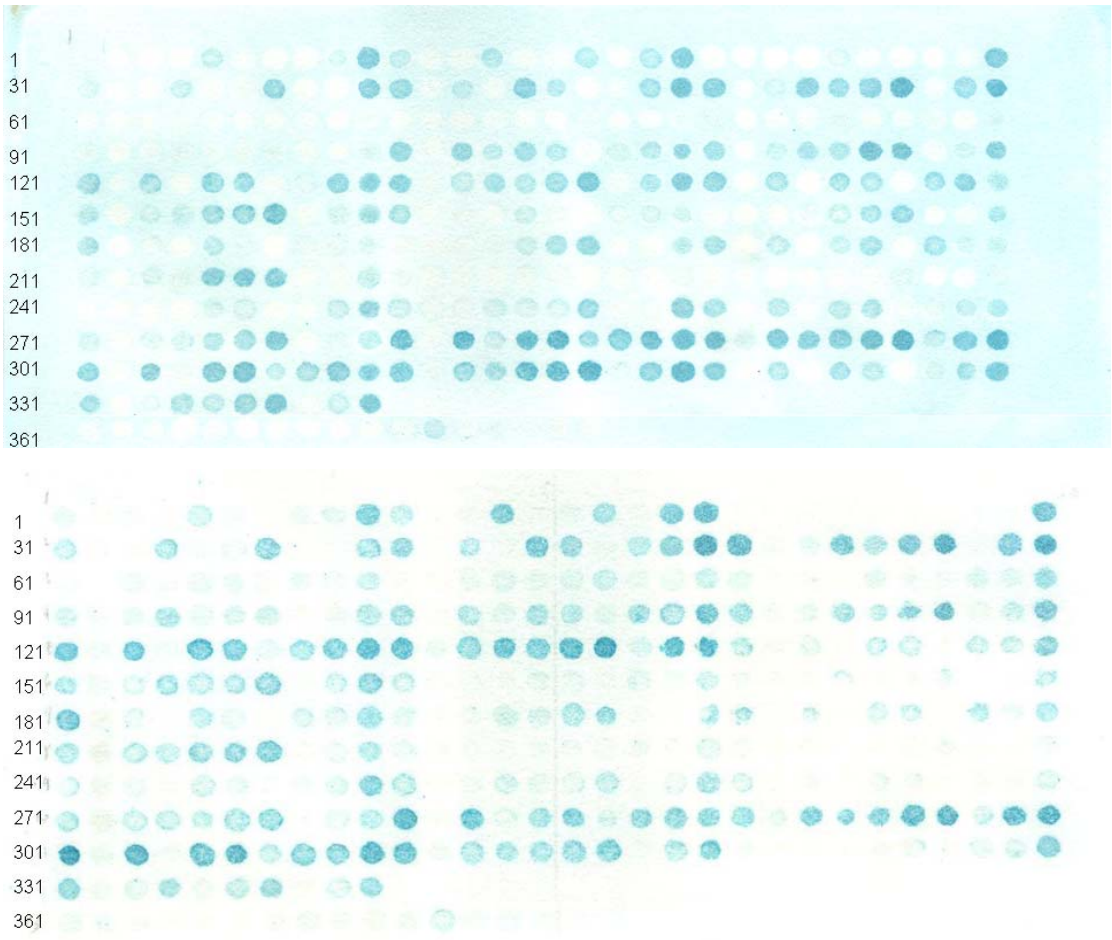


Figure 3.10 Screening of enzymatic prenylation of an RAGCVa_2X library. Each membrane was prenylated with OPP-Far-alkyne by PFTase, clicked with biotin-azide and then visualized

by SP-AP in BCIP solution. Screenings from two separate membranes are shown.

Spot identity:

1-20: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)E

21-40: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)D

41-60: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)N

61-80: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)R

81-100: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)K

101-120: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)H

121-140: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)A

141-160: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)V

161-180: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)I

181-200: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)L

201-220: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)F

221-240: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)W

241-260: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)G

261-280: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)C

281-300: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)M

301-320: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)S

321-340: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)T

361-377: SVI(E,D,N,R,K,H,A,V,I,L,F,W,G,C,M,S,T)

X \ a	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V
E	7	2	3	2	22	4	3	4	15	82	30	6	7	50	8	8	35	5	43	74
D	3	2	2	1	9	3	2	3	4	77	17	3	3	27	4	5	59	4	3	60
N	57	6	34	9	80	49	8	19	56	100	84	5	15	70	48	70	104	5	60	73
R	3	2	4	4	6	5	5	7	4	12	5	6	9	14	11	12	18	8	6	16
K	11	4	5	4	11	6	6	10	8	29	13	5	8	25	8	12	18	5	6	25
H	59	6	53	25	53	37	11	33	42	63	48	5	16	44	31	70	76	6	19	68
A	61	6	58	5	74	64	8	23	68	81	78	9	44	66	53	77	104	13	61	78
V	51	5	32	2	35	32	3	37	42	42	33	7	18	35	43	50	79	4	17	55
I	35	5	9	8	23	12	7	14	12	18	6	5	3	8	12	25	37	2	3	25
L	51	3	10	3	29	9	3	12	12	28	9	6	8	18	26	48	59	6	4	30
F	41	5	23	2	19	33	2	37	10	22	20	4	21	20	63	70	79	4	6	29
W	10	5	7	7	10	8	5	7	4	12	5	4	2	2	3	5	11	3	3	7
G	6	3	5	3	19	14	4	5	22	63	29	7	9	23	18	17	50	6	8	63
C	34	3	23	3	25	30	6	10	18	26	23	4	17	21	31	48	70	3	11	32
M	95	5	95	22	77	91	37	68	78	81	78	16	71	57	67	94	103	25	81	99
S	79	5	80	6	83	89	27	59	73	84	79	8	51	47	76	87	99	10	55	89
T	43	2	22	1	28	22	2	12	29	73	50	4	11	40	23	47	70	5	20	68

Figure 3.11 Evaluation of the extent of farnesylation of a RAGCVa_2X library of peptides. After synthesis, the peptides were subjected to enzymatic prenylation with an alkyne-containing analogue of FPP followed by click reaction with biotin-azide and subsequent reaction with SP-AP using a chromogenic substrate. Color intensity was quantified by Image J software. For comparison, the intensity was normalized relative to that observed with Asn at the X position. The library was synthesized and screened two times and

the average color intensities were color coded. Intensities below 33% are shown in white, tintensities between 34 to 65% are shown in yellow and tintensities above 65% are shown in red.

Two general trends emerge from inspection of this data. First, analysis clearly indicates that PFTase from *S. cerevisiae* prefers M or S residues at the X position. This can be concluded from simple inspection of the number of high intensity (red) signals in their respective rows (Figure 3.11) or more quantitatively by comparison of the summation of all intensities in a given row (Figure 3.12). Such an analysis reveals that N, and A, and to a lesser extent, H V, F and T are also tolerated at the X position. A similar analysis indicates a preference for I, S, T and V residues at the a_2 position with a number of others occurring less frequently. In addition, at the level of individual sequence analysis, a survey of all of the peptide sequences giving rise to the strongest signals (53 total in red), 17 have previously been reported to be substrates for the mammalian PFTase;[112] only two sequences, CVID and CVIE, that are not good substrates appeared as strong signals (as noted above). Additionally, 38 out of the 53 sequences are predicted to be substrates for the related human PFTase using the computer algorithm PrenPS[116] (Figure 3.13). Moreover, of the 15 non-substrate sequences predicted by PrenPS, two have been evaluated *in vitro*; both of those have been shown to be effective substrates. Hence, we conclude that the

strategy described here using C-terminal peptide libraries generated by robotic spot synthesis can be used to identify *bona fide* substrate sequences.

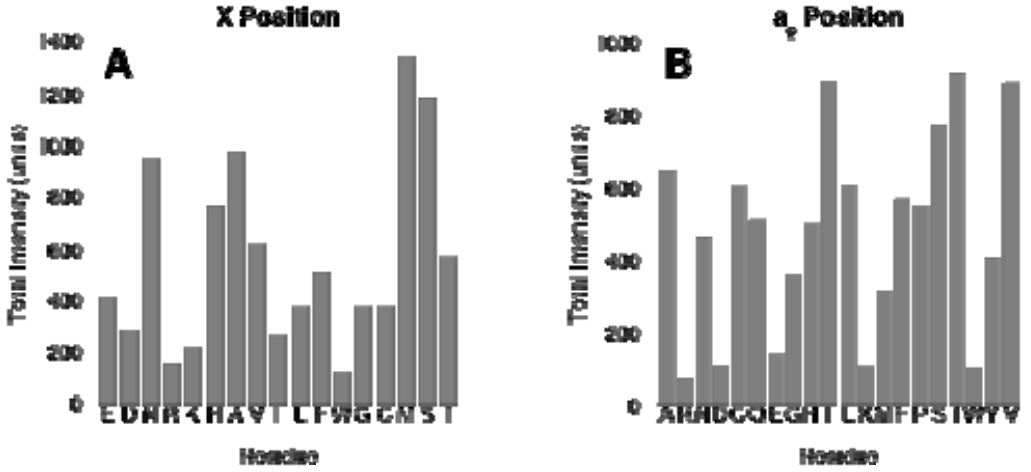


Figure 3.12 Analysis of preferred residues at the X position and a_2 position for a Ca_1a_2X-box obtained from screening of a 340-member library of peptides. Left (A): Preferences for the X position. Right (B): Preferences for the a_2 position. These preferences were determined by summing the intensities obtained for a single residue at the X or a2 position across all possible combinations at the adjacent position. For example, to calculate the preference for E at the X position, all intensity values present in the "E" row in Fig. 3 were summed (7+2+3+2+22 etc). To calculate the preference for A at the a_2 position, all intensity values present in the "A" column in Fig. 3 were summed (7+3+57+3+11 etc).

sequence	Screening value	PrePS		Kinetic assay[6]	sequence	Screening value	PrePS		Kinetic assay[6]
CVAM	95	-3.2	(-)	NA	CVFA	66	-3.6	(-)	NA
CVAS	79	-3.6	(-)	NA	CVPM	67	-0.4	(+)	NA
CVNM	95	-3.3	(-)	NA	CVPS	76	-0.7	(+)	NA
CVNS	80	-3.6	(-)	NA	CVSN	70	-4.0	(-)	NA
CVCN	80	-3.6	(-)	NA	CVSH	70	-2.0	(+)	NA
CVCA	74	-0.9	(+)	Sub	CVSA	77	-1.3	(+)	NA
CVCM	77	-0.5	(+)	NA	CVSF	70	-1.6	(+)	NA
CVCS	83	-0.9	(+)	NA	CVSM	94	-0.9	(+)	NA
CVQM	91	0.4	(++)	MTO	CVSS	87	-1.3	(+)	STO
CVQS	89	0.2	(++)	NA	CVTN	104	-2.7	(-)	NA
CVGM	68	-3.2	(-)	NA	CVTH	76	-0.7	(+)	NA
CVHA	68	-3.7	(-)	NA	CVTA	104	-0.0	(+)	Sub
CVHM	78	-3.7	(-)	NA	CVTV	79	-0.7	(+)	NA
CVHS	73	-3.7	(-)	MTO	CVTF	79	-0.4	(+)	NA
CVIE	82	-1.3	(+)	Sub	CVTC	70	0.2	(++)	NA
CVID	77	-1.9	(+)	NA	CVTM	103	0.4	(++)	NA
CVIN	100	-1.6	(+)	Sub	CVTS	99	0.0	(++)	NA
CVIA	81	1.1	(++)	Sub	CVTT	70	-0.7	(+)	MTO
CVIM	81	1.4	(++)	Sub	CVYM	81	-4.1	(-)	NA
CVIS	84	1.1	(++)	Sub	CVVE	74	-1.7	(+)	NA
CVIT	73	0.5	(++)	Sub	CVVN	73	-2.1	(-)	MTO
CVLN	84	-2.4	(-)	NA	CVVH	68	-0.2	(+)	NA
CVLA	78	0.3	(++)	Sub, MTO	CVVA	78	0.6	(++)	Sub
CVLM	78	0.6	(++)	NA	CVVM	99	0.9	(++)	Sub
CVLS	79	0.3	(++)	Sub	CVVS	89	0.6	(++)	NA
CVMM	71	-0.3	(+)	NA	CVVT	68	-0.1	(+)	MTO
CVFN	70	-6.5	(--)	NA					

Figure 3.13 Comparison of our screening result with Prenylation Prediction Suite and the kinetic assay performed before. Only peptides sequences giving the high intensity from the screening are included in this table. For the Prenylation Prediction Suite, the upstream sequence of the Ca_1a_2X is GGLLBBQFRAG. For the kinetic assay, NA means sequences not measured, Sub means Ca_1a_2X sequences from known farnesylated proteins, MTO means sequences exhibit multiple-turnover activity and STO means sequences exhibit single-turnover activity.

3.3 Conclusion

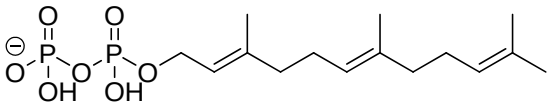
From a biological perspective, it should be noted that in previous work, CaaX-box sequences where X is A, M, Q or S have been shown to be prenylated by PFTase from *S. cerevisiae*;[109] however, N and T have not been reported as substrates for the yeast enzyme. These results highlight the utility this library approach for exploring the sequence specificity of PFTase and other enzymes involved in the subsequent processing of prenylated proteins. Efforts to employ such arrays to probe the specificity of the prenyltransferases and other prenylated protein processing enzymes are introduced in the next chapter.

3.4 Materials and methods

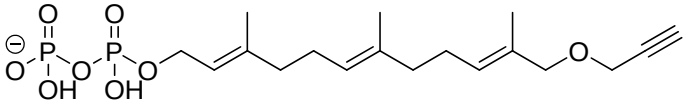
(a) General

All synthetic reactions were carried out at 25^{o}C and stirred magnetically unless otherwise noted. TLC was performed on precoated (250 mm) silica gel 60 F-254 plates (Merck). Plates were visualized by staining with $KMnO_4$ or hand-held UV lamp. Flash chromatography silica gel (60–200 mesh, 75–250 μm) was obtained from Mallinckrodt Inc. Deuterated NMR solvents were purchased from Cambridge Isotope Laboratories, Inc. Biotin-azide was purchased from ChemPep. All amino

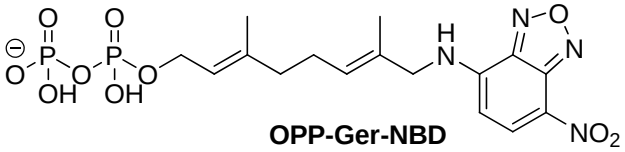
acids and peptide coupling reagents were purchased from NovaBiochem. Other general reagents, including salts and reagents, were purchased from Sigma-Aldrich. ^{1}H NMR spectra were obtained at 300 MHz; ^{31}P NMR spectra were obtained at 121 MHz. All NMR spectra were acquired on Varian instruments at 25°C. Chemical shifts are reported in ppm and *J* values are in Hz. SPOT synthesizer is ResPep SL obtained from INTAVIS Bioanalytical Instruments. Fluorescence assay data were obtained by using a Varian Cary Eclipse Fluorescence Spectrophotometer. Analytical HPLC was performed on a Beckman model 125/166 instrument, equipped with a diode array UV detector, ABI Analytical Spectroflow 980 fluorescence detector, and a Varian C_{18} column (Microsorb-MV, 5 μm, 4.6x250 mm). Preparative HPLC separations were performed by using a Beckman model 127/166 instrument, equipped with a UV detector and a Phenomenex C_{18} column (Luna, 10 μm, 10x250 mm). Fluorescence scanning was done by Storm 840 Phosphor-Chemifluorescence Workstation (excitation 450 nm & emission 540-560 nm). MS spectra for small molecules were obtained on a Bruker BioTOF II instrument. MS spectra for peptide libraries were obtained on an AB Sciex 4800 MALD TOF/TOF or Bruker Biflex III instrument. Yeast PFTase[117], Ds-GCVIA[1] and OPP-Far-alkyne[115] were prepared as previously described. OPP-Ger-NBD was prepared using a reported reaction procedure.[3]



FPP



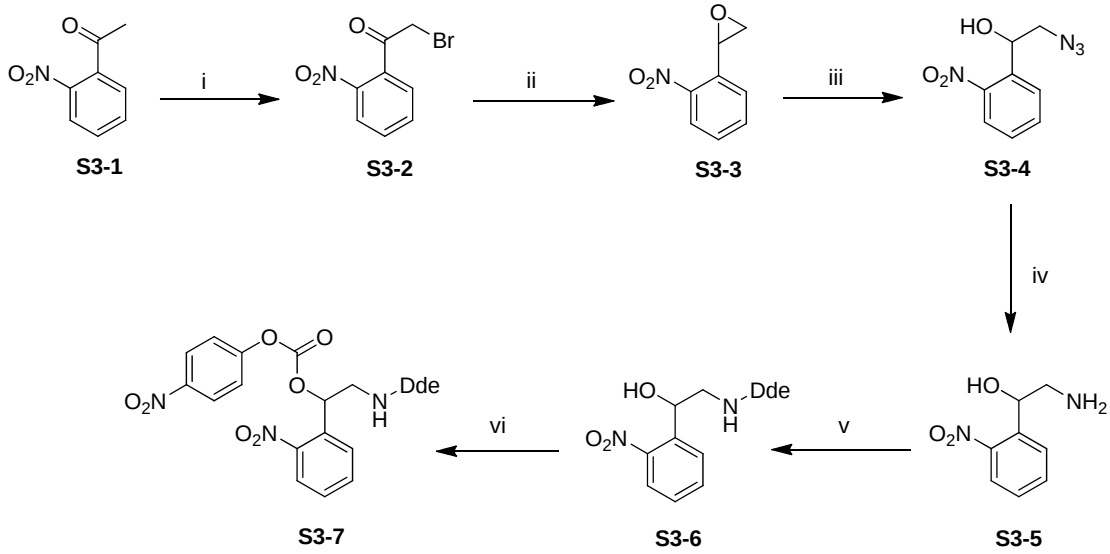
OPP-Far-alkyne



OPP-Ger-NBD

Analogues used to screen peptide libraries for PFTase activity.

(b) Synthesis of the photocleavable linker



Scheme S3.1 Synthesis of the photocleavable linker used for library synthesis. Reagents and conditions: (i) Br_2, 1,4-dioxane, room temp, 3 h; (ii) $NaBH_4$, H_2O/1,4-dioxane, 0 ^oC to RT, 1.5 h, then $NaOH_{(aq)}$, RT, 15 min; (iii) NaN_3, NH_4Cl, EtOH/H_2O, reflux, 12 h; (iv) PPh_3, THF, 0 ^oC to RT, 2 h, then H_2O, RT, 12 h; (v) Dde-OH, CH_2Cl_2, RT, 18 h; (vi) p-nitrophenyl chloroformate, pyridine, CH_2Cl_2, RT, 350 min.

O-nitrophenacyl bromide (S3-2)

Steps (i) and (ii) were performed according to a previously published procedure.[4] A solution of Br_2 (4.5 mL, 87.6 mmol) in dioxane (80 mL) was added dropwise over a 30 min period to a solution of o-nitroacetophenone **(S3-1)** (14.44 g, 87.4 mmol) in dioxane (40 mL). After the addition was completed, the mixture was stirred for 2.5 h. The solvent and HBr formed were removed by rotary evaporation followed by the addition of H_2O (60 mL). The solution was extracted with Et_2O (3×50 mL) and the combined organic layers were dried with $MgSO_4$ and concentrated by rotary evaporation to yield **S3-2** as a brown oil. Crude compound (**2**) was used without further purification. ^{1}H NMR ($CDCl_3$): δ 4.30 (s, 2H), 7.51 (dd, *J* = 1.5, 7.2, 1H), 7.68 (ddd, *J* = 1.8, 7.8, 8.4, 1H), 7.80 (ddd, *J* = 1.2, 7.5, 7.5 1H), 8.24 (dd, *J* = 1.2, 8.4, 1H).

O-nitroepoxystyrene (S3-3)

A suspension composed of $NaBH_4$ (3.31 g, 87.4 mmol) and H_2O (30 mL) was added dropwise to a solution of o-nitrophenacyl bromide (**S3-2**) (prepared in previous step) in dioxane (30 mL) over a 40 min period at 0 °C with stirring. The mixture was stirred at 0 ^{o}C for 40 min and then at RT for 70 min. An aqueous solution of NaOH (4.14 g / 30 mL) was added to the mixture. The mixture was stirred at RT for 15 min.

H_2O (40 mL) was then added and the solution was extracted with Et_2O (3×50 mL). The combined organic layers were dried with $MgSO_4$ and concentrated by rotary evaporation. The resulting oil was purified by silica gel column chromatography (1:60 Et_2O/hexanes containing 2% Et_3N) to yield **S3-3** as a yellow solid (8.63 g, 59.8% yield). ^{1}H NMR ($CDCl_3$): δ 2.68 (dd, *J* = 2.7, 5.4, 1H), 3.30 (dd, *J* = 4.5, 5.7, 1H), 4.49 (dd, *J* = 2.7, 4.5, 1H), 7.48 (ddd, *J* = 1.8, 6.9, 6.9, 1H), 7.65 (m, 2H), 8.16 (dd, *J* = 1.2, 8.4, 1H).

2-azido-1-(2-nitrophenyl)ethanol (S3-4)

To a mixture of compound (**S3-3**) (8.63 g, 52.3 mmol) in EtOH (70 mL) was added NaN_3 (4.08 g, 62.8 mmol), NH_4Cl (2.87 g, 53.7 mmol) and then water (70 mL). The mixture was stirred at reflux for 16 h. EtOH was removed by rotary evaporation. H_2O (40 mL) was added and the solution was extracted with Et_2O (3×50 mL). The combined organic layers were dried with $MgSO_4$ and concentrated by rotary evaporation. The resulting oil was purified by silica gel column chromatography (1:3 Et_2O/hexanes) to yield **S3-4** as a yellow oil (4.41 g, 40.5% yield). ^{1}H NMR ($CDCl_3$): δ 3.46 (dd, *J* = 7.5, 12.3, 1H), 3.76 (dd, *J* = 3.0, 12.3, 1H), 5.51 (dd, *J* = 3.3, 7.2, 1H), 7.50 (ddd, *J* = 1.5, 8.4, 8.4, 1H), 7.71 (ddd, *J* = 1.2, 7.5, 7.5, 1H), 7.92 (dd, *J* = 1.5, 7.8, 1H), 8.02 (dd, *J* = 1.2, 8.4, 1H).

2-amino-1-(2-nitrophenyl)ethanol (S3-5)

Compound **S3-4** (4.41 g, 21.2 mmol) was dissolved in dry THF (35 mL). The mixture was placed under an N_2 (g) atmosphere and cooled to 0 ^oC. PPh_3 (5.85 g, 22.3 mmol) was added to this mixture, stirred at 0 ^oC for 10 min and then RT for 110 min. H_2O (3 mL) was added to this mixture which was then stirred at RT for 16 h. Et_2O (50 mL) was added and the solution was extracted with 1M HCl (2×35 mL). The aqueous layers were combined, adjusted to a basic pH with $NaOH_{(s)}$ and extracted with Et_2O (3×40 mL). The combined organic layers were dried with $MgSO_4$ and concentrated by rotary evaporation to yield compound **S3-5** as a yellow solid that was used without further purification. ^1H NMR (d_6-acetone): δ3.79 (dd, J = 7.2, 12.3, 1H), 5.29 (dd, J = 5.7, 6.6, 1H), 7.54 (ddd, J = 1.5, 7.8, 8.7, 1H), 7.72 (ddd, J = 1.2, 7.2, 8.7, 1H), 7.89 (dd, J = 1.2, 7.8, 1H), 8.02 (dd, J = 1.2, 8.1, 1H).

2-(1-(2-hydroxy-2-(2-nitrophenyl)ethylamino)ethylidene)-5,5-dimethylcyclohexane-1,3-dione (S3-6)

Preparation of Dde-OH: Dimedone (5.0 g, 0.0357 mol) was dissolved in DMF (75 mL) with HOAc (2.0 mL, 34.9 mmol), DCC (7.38 g, 35.8 mol), and DMAP (4.36 g, 35.7 mmol) and allowed to react for 16 h. Precipitating dicyclohexylurea (DCU)

was removed by filtration, and the solvent was evaporated in vacuo. After dissolution in EtOAc (50 mL) and washed with 1M HCl (2×50 mL) and subsequently with H_2O (2×50 mL), the organic phase was dried over $MgSO_4$, filtered and concentrated to give a yellowish liquid that slowly solidified upon standing. ^{1}H NMR ($CDCl_3$): δ 1.08 (s, 6H), 2.36 (s, 2H), 2.53 (s, 2H), 2.61 (s, 3H). ESI-MS [M-H$^+$]: m/z calc. for $C_{10}H_{13}O_3^-$: 181.0870; found:181.0852.

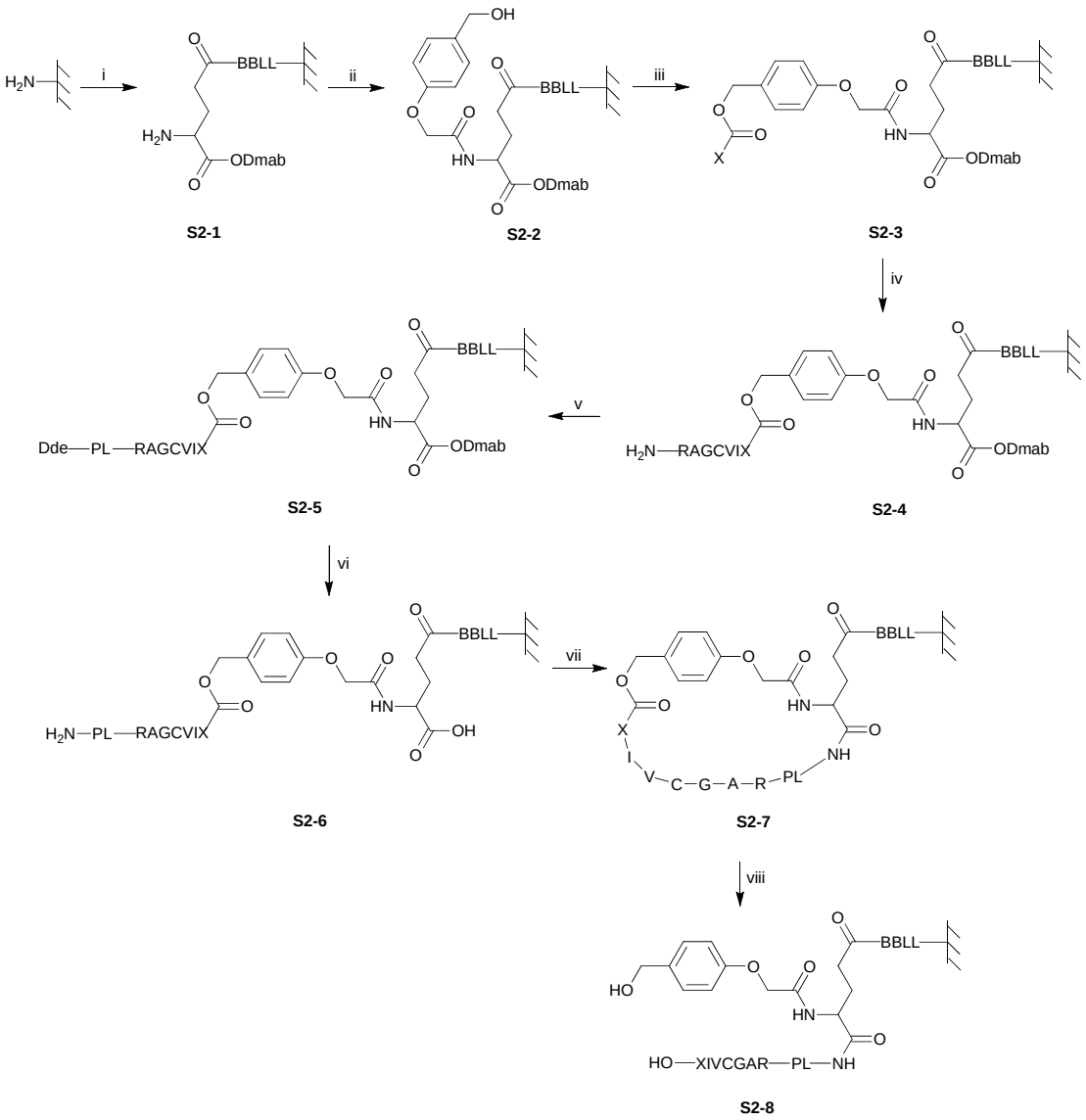
Protection with Dde-OH: Compound **S3-5** (1.48 g, 8.15 mmol) was stirred with Dde-OH (1.45 g, 7.96 mmol) in 20 mL CH_2Cl_2 at RT for 16 h. Some CH_2Cl_2 was removed by rotary evaporation and the resulting mixture was purified by silica gel column chromatography (3:1 Et_2O/hexanes) to yield **S3-6** as a yellow solid (2.06 g, 73.1 %). ^{1}H NMR ($CDCl_3$): δ 0.97 (s, 6H), 2.27 (s, 4H), 2.53 (s, 3H), 3.98 (m, 2H), 5.55 (dd, *J* = 3.3, 7.5, 1H), 7.54 (ddd, *J* = 7.5, 7.5, 1H), 7.72 (dd, *J* = 7.8, 7.8, 1H), 8.02 (m, 2H). ESI-MS [M+Na$^+$]: m/z calc. for $C_{18}H_{22}N_2O_5Na^+$: 369.1421; found: 369.1480.

2-(1-(4,4-dimethyl-2,6-dioxocyclohexylidene)ethylamino)-1-(2-nitrophenyl)ethyl 4-nitrophenyl carbonate (S3-7)

To a solution of compound **S3-6** (0.454 g, 1.31 mol) in CH_2Cl_2 (5 ml) was added 4-nitrophenyl chloroformate (0.394 g, 1.95 mmol) and pyridine (0.19 ml, 2.35 mmol).

The mixture was stirred at room temperature for 350 min. Some CH_2Cl_2 was removed by rotary evaporation and the resulting mixture was purified by silica gel column chromatography (3:1 Et_2O/hexanes) to yield **S3-7** as a white solid (0.487 g, 72.7 %). ^{1}H NMR ($CDCl_3$): δ 1.047 (s, 6H), 2.36 (s, 2H), 2.43 (s, 2H) 2.57 (s, 3H), 3.97 (dd, *J* = 5.7, 14.4, 1H), 4.22 (dd, *J* = 3.3, 9.6, 1H), 6.50 (d, *J* = 4.2, 1H), 7.38 (d, *J* = 9.6, 2H), 7.61 (dd, *J* = 8.1, 8.1, 1H), 7.88 (m, 2H), 8.14 (d, *J* = 8.1, 1H), 8.25 (d, *J* = 7.8, 2H). ESI-MS [M+H$^+$]: m/z calc. for $C_{25}H_{26}N_3O_9^+$: 512.1664; found: .512.1591.

(c) Synthesis of inverted peptides



Scheme S3.2 Synthesis of inverted peptides. Reagents and conditions: (i) standard DIC coupling (2×), then, capping, then 20% piperidine, then BPB staining; (ii) standard DIC coupling (2×); (iii) 0.5 M Fmoc-Aa and 0.5 M 6-Cl-HOBt in DMF/CH_2Cl_2, 0.5 M DIC in CH_2Cl_2, 0.2 M DMAP in CH_2Cl_2, (6-8×), then, capping, then 20% piperidine, then BPB staining; (iv) standard DIC coupling (2×), then, capping, then 20% piperidine, then BPB staining; (v) 0.5 M photocleavable linker, 0.5 M Et_3N in DMF (3-4×); (vi) 2% NH_2NH_2; (vii) 0.05 M BOP, 0.05 M 6-Cl-HOBt and 0.1 M DIEA in DMF (2×); (viii) modified reagent K.

Amino acid residues used in the library synthesis

Abbreviation	Chemical name	Abbreviation	Chemical name
B	Fmoc-β-Ala-OH	K	Fmoc-Lys(Boc)-OH
L	Fmoc-Leu-OH	H	Fmoc-His(Trt)-OH
R	Fmoc-Arg(Pbf)-OH	P	Fmoc-Pro-OH
A	Fmoc-Ala-OH	F	Fmoc-Phe-OH
G	Fmoc-Gly-OH	Y	Fmoc-Tyr(tBu)-OH
C	Fmoc-Cys(Trt)-OH	W	Fmoc-Trp(Boc)-OH
V	Fmoc-Val-OH	M	Fmoc-Met-OH
I	Fmoc-Ile-OH	S	Fmoc-Ser(tBu)-OH
Q	Fmoc-Gln(Trt)-OH	T	Fmoc-Thr(tBu)-OH
D	Fmoc-Asp(OtBu)-OH	E	Fmoc-Glu(OtBu)-OH
N	Fmoc-Asn(Trt)-OH		Fmoc-Glu-ODmab (For cyclization)

Manual synthesis:

The inverted peptides were synthesized on amino-functionalized cellulose membranes (0.40 mmol/g). All of the manipulations were performed at RT unless otherwise noted. The peptides were synthesized by spotting 1 μL of 0.5 M Fmoc-amino acid and 1 μL of 0.5 M DIC in DMF. The coupling reaction was typically allowed to proceed for 30 min twice. The membrane was incubated with 2% (v/v) Ac_2O in DMF (capping solution A) for 5 min, followed by 2% (v/v) acetic anhydride, 2% (v/v) DIEA in DMF (capping solution B) for 15 min without shaking. The membrane was then washed with DMF (3×). The Fmoc groups were removed by treatment twice with 20% piperidine in DMF (5+5 min) followed by washing of the membrane with DMF (4×) and MeOH (2×). The membrane was then stained with BPB in MeOH (40 mg/L) to visualize the spot locations. The membrane was then

washed with MeOH (4×) and Et_2O (2×).

Next, HMPA was coupled to the growing peptide chain using the above procedure. The membrane was washed with MeOH (4×), diethyl ether (2×) without capping.

Next, esterification of the links which will be the C-terminal residue in the final product was accomplished by spotting sequentially 1 μL of 0.5 M Fmoc-amino acid, 6-Cl-HOBt in DMF/CH_2Cl_2 (v/v = 50:50), 1 μL of 0.5 M DIC in CH_2Cl_2 and 1 μL of 0.2 M DMAP in CH_2Cl_2. Each coupling reaction was typically allowed to proceed for 30 min. The process was repeated for six to eight times. The membrane was then incubated with capping solution A for 5 min followed by capping solution B for 15 min without shaking. The membrane was washed with DMF (3×) and the Fmoc group was removed by treatment twice with 20% piperidine in DMF (5+5 min). The membrane was washed with DMF (4×), MeOH (2×). The membrane was stained with BPB in MeOH (40 mg/L) to indicate spot locations. The membrane was washed with MeOH (4×) followed by Et_2O (2×).

To add the photocleavable linker, 1 μL of 0.5 M photocleavable linker solution and 1 μL of 0.5 M Et_3N in DMF were spotted onto the membrane. The coupling reaction was typically repeated three to four times with each time to proceed for 30 min.

Next, the Dmab and Dde protecting groups were removed by treatment with 2% hydrazine solution in DMF (3×10 min) followed by washing with DMF and MeOH.

For peptide cyclization, a solution of (BOP)/6-Cl-HOBt/DIEA (0.05 M, 0.05 M, 0.1 M respectively) in DMF was mixed with the membrane. This reaction was performed twice, the first time for 2 h and the second time for 16 h. The membrane was washed with DMF (3×), MeOH (2×) and Et_2O (2×).

The membrane was treated with a modified reagent K (6.5% phenol, 5% H_2O, 5% thioanisole, 2.5% ethanedithiol, 1% anisole, and 1% triisopropylsilane in TFA) for 2 h for ester cleavage (ring opening) and side-chain deprotection.

SPOT Synthesizer:

The peptides were synthesized by ResPep SL SPOT-robot (INTAVIS Bioanalytical Instruments). The synthesis started with spot definition by a standard protocol followed by the coupling of a solution of 0.11 μL of Fmoc-amino acid (0.5 M), 0.055 μL of DIC (1.1 M) and 0.055 μL of 6-Cl-HOBt (1.2 M) in DMF (double coupling, 10 and 15 min reaction each). The membrane was acetylated with acetic anhydride in DMF (2%), washed with DMF (3×) and EtOH (3×) and finally air-dried. Fmoc was removed by 20% piperidine in DMF (3×). After HMPA was coupled to the growing peptide chain using the above procedure, the membrane was washed with DMF (3×), EtOH (3×) without capping.

For ester-bond formation (step iii), 0.22 μL of solution Fmoc-amino acid (0.4M) activated with 1,1'-carbonyldiimidazole (CDI, 3 equiv.) in DMF were spotted on the membrane (4× coupling, 15 min reaction time each).

To add the photocleavable linker, 0.22 μL of 0.5 M photocleavable linker solution mixed with Et_3N (2 equiv.) in DMF were spotted onto the membrane (4× coupling, 15 min reaction time each).

The procedure of peptide cyclization, side-chain cleavage and other steps was the same as manual synthesis.

(d) MALDI MS analysis of peptides obtained via SPOT synthesis before and after enzymatic prenylation

Each spot membrane was incubated with 500 μL CH_3CN/H_2O (v/v = 5/95) in a microcentrifuge tube. Peptide release was accomplished by shining UV light (λ = 365 nm) on the side of the tube for 1 h. After photolysis, the solvent was transferred to a new microcentrifuge tube. The membrane was washed with 500 μL CH_3CN and then 500 μL MeOH. The combined solution was concentrated *in vacuo*. The resulting residue was dissolved in 70 μL CH_3CN/H_2O (v/v = 5/95) with 0.1% TFA. In a typical MALDI experiment, 0.7 μL of peptide solution was mixed with 0.7 μL CHCA matrix (10 mg/mL; CH_3CN/H_2O (v/v = 50/50) with 0.1% TFA) on the MALDI target and air

dried.

(e) Screening for PFTase specificity with OPP-Far-alkyne

Spots from manual synthesis:

Spots from the membranes produced above were excised and incubated with 155 mM DTT for 1 h and then washed with H_2O (1×). Typical conditions for PFTase reactions were 5.1 mM DTT, 5 mM $MgCl_2$, 50 μM $ZnCl_2$, 200 mM Tris (pH=7.5), 1 mg/mL BSA, 260 μM OPP-Far-alkyne, 0.09 mg/mL PFTase. The membranes were incubated in the solution for 30 min before adding OPP-Far-allkyne and PFTase. The enzymatic reactions were typically allowed to proceed for 4 h followed by washing with 25 mM $NH_4HCO_{3(aq)}$ (4×), CH_3CN (4×) after which the membranes were transferred to new 1.5 mL microcentrifuge tubes.

Typical conditions for click reactions were 0.1 mM biotin-azide, 1 mM TCEP, 0.2 mM TBTA, 1 mM $CuSO_4$ in PBS buffer 1 (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 2mM KH_2PO_4, pH=7.4). The reaction was typically allowed to proceed for 16 h. After the reactions, the membranes were washed with H_2O (4×), DMF (4×) and placed in new tubes.

The membranes were incubated with 5 % milk in PBS buffer 2 (100 mM KH_2PO_4, 500 mM NaCl, pH=6.5) for 1 h and then incubated with SP-AP in 5% milk

PBS buffer 2 (v/v = 1/800) for 30 min. After the reactions, the membranes were washed with PBS buffer 2 (7×) and placed in new tubes. Upon addition of 750 µL of buffer E (5 mM $MgCl_2$, 20 µM $ZnCl_2$, 100 mM NaCl, 30 mM Tris, pH=8.5) and 50 µL of 5 mg/mL BCIP solution, turquoise color developed on positive spots after 1 h. The staining was terminated by adding 200 uL of 10 M HCl.

Spots from SPOT synthesizer:

The whole membrane was incubated with 155 mM DTT for 1 h and then washed with H_2O (1×). Typical conditions for PFTase reactions were 5.1 mM DTT, 5 mM $MgCl_2$, 50 µM $ZnCl_2$, 100 mM Tris (pH=7.5), 1 mg/mL BSA, ~100 µM OPP-Far-alkyne, 0.045 mg/mL PFTase. The enzymatic reactions were typically allowed to proceed for 4.5 h followed by washing with 25 mM $NH_4HCO_{3(aq)}$ (3×), CH_3CN (3×), and H_2O (3×).

Typical conditions for click reactions were 0.1 mM biotin-azide, 1 mM TCEP, 0.2 mM TBTA, 1 mM $CuSO_4$ in PBS buffer 1. The reaction was typically allowed to proceed for 16 h. After the reactions, the membranes were washed with H_2O (3×), DMF (3×), and H_2O (3×).

The membranes were incubated with 5 % milk in PBS buffer 2 for 1 h and then incubated with SP-AP in 5% milk PBS buffer 2 (v/v = 1/800) for 30 min. After the reactions, the membranes were washed with H_2O (9×). Upon addition of BCIP

solution in buffer E (0.313 mg/mL), turquoise color developed on positive spots after 1.5 h. The staining was terminated by washing with H_2O (2×) and CH_3CN (1×).

The quantification of color intensity was performed using Image J software. The membrane was scanned and saved. The file was opened by Image J. Under Image>Type, 8-bit was selected to convert the image to grayscale. Under the menu Process>Subtract Background, the rolling ball radius was set to 50. Under Analyze>Set Measurements, the Area, Mean Gray Value, and Integrated Density were selected. Under Analyze>Set Scale, "pixels" in the box next to Unit of length were selected. Under Edit>Invert, the colors on the image were inverted, a line was drawn around the image boundary and the measurement of the selected are was then performed using the "m" key.

(f) Kinetic analysis of individual peptides for substrate verification

The rates of farnesylation of individual purified peptides by PFTase were determined using the time-dependent increase in fluorescence (λ_{ex} 340 nm, λ_{em} 510 nm) upon prenylation of a dansylated form of the peptide. Assays were performed with 2.4 μM dansylated peptide, 70 nM PFTase, 10 μM OPP-Far-alkyne, 200 mM Tris, pH 7.5, 5 mM DTT, 5 mM $MgCl_2$, 50 μM $ZnCl_2$, and 0.040 % (*w/v*)

n-dodecyl-*β*-D-maltoside at 25 °C. Peptides were incubated in reaction buffer for 5 min prior to initiation by the addition of PFTase. Fluorescence was measured as a function of time to define both the initial linear velocity and the reaction end point. The total fluorescence change observed upon reaction completion was divided by the initial concentration of the peptide substrate in a given reaction to yield a conversion from fluorescence units to product concentration (Amp_{conv}). The linear initial rate, V, in fluorescence intensity per minute, was then converted to a velocity (μM product produced per minute) by dividing V with Amp_{conv}.

Chapter 4: Rapid Analysis of Protein Farnesyltransferase Substrate Specificity Using Peptide Libraries and Isoprenoid Diphosphate Analogues

4.1 Introduction

In the last chapter and previous work, we reported the screening of a library of peptides for catalytic activity using *S. cerevisiae* PFTase (yPFTase).[39] A similar strategy was used here for peptide synthesis and subsequent evaluation. Here we applied the SPOT synthesis method[118] to study the specificity of *R. norvegicus* PFTase (rPFTase) and investigate the interplay between peptides and isoprenoid substrates of varying length (Figure 4.1) and the specificity of PFTases from different organisms.

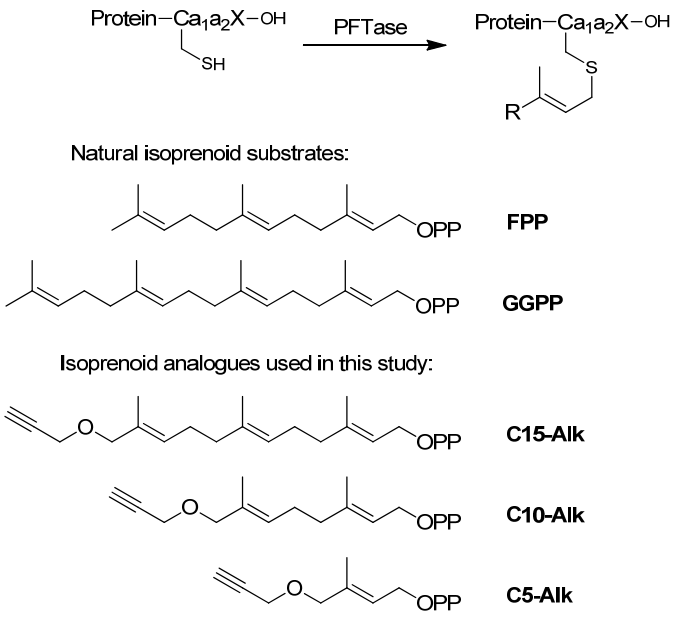


Figure 4.1 The post-translational prenylation reactions catalyzed by PFTase and PGGTase I and the probes used here to investigate PFTase. The ‘‘R’’ group can be a farnesyl or geranylgeranyl group in cells.

4.2 Peptide library design, synthesis and screening

An automated SPOT synthesizer was used to create two kinds of peptide libraries (a 19×20 CVa_2X library and a 19×20 CCa_2X library with the X being one of the 20 proteogenic amino acids except P and a_2 being one of the 20 proteogenic amino acids). Because peptides are chemically synthesized in a C- to N-terminal direction, we adopted a "peptide inversion" strategy to prepare peptide libraries with free C-termini.[31,32,44] In this approach, synthetic peptides are cyclized between their N-terminus and an internal carboxyl group installed on a bifunctional linker followed by acidolytic global deprotection and ester cleavage to yield resin-bound peptides with free C-termini (Figure 4.2a). To confirm the production of the desired synthetic peptides, a photocleavable linker was incorporated N-terminal to the Ca_1a_2X sequence so that at the end of the synthesis, individual peptide could be released from the membrane by UV irradiation and analyzed by MALDI. Following synthesis, each membrane was subjected to PFTase-catalyzed prenylation with an alkyne-containing FPP analogue followed by derivatization with biotin-azide via copper-catalyzed azide-alkyne cycloaddition (CuAAC). Those peptides that were prenylated by PFTase were conjugated to biotin at this step. The membrane was then subjected to an enzyme-linked assay involving streptavidin-alkaline phosphatase (SA-AP) and the chromogenic substrate 5-bromo-4-chloro-3-indolyl phosphate (BCIP). Spots

containing prenylated peptides appear turquoise colored, whereas spots where the prenylation reaction was inefficient remain colorless (Figure 4.2b).

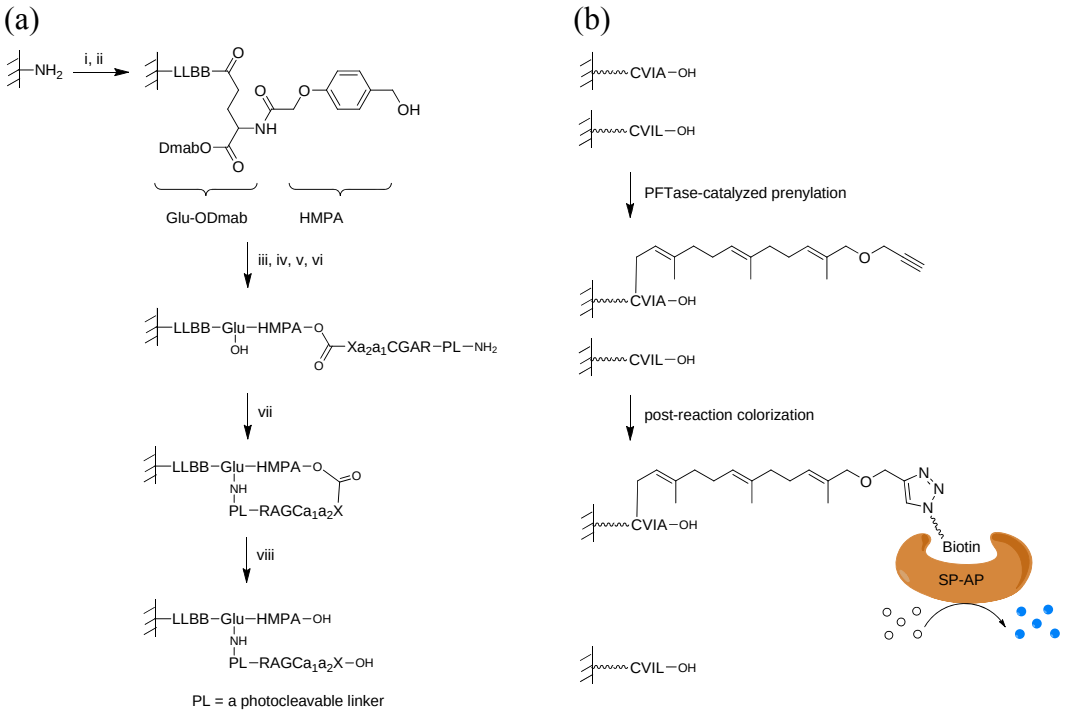


Figure 4.2 Strategy for the synthesis of C-terminal Ca_1a_2X-box peptide libraries and their subsequent use to explore the specificity of PFTase. (a) Synthesis of C-terminal peptides. Reagents and conditions: (i) standard DIC coupling of Fmoc-Aa (2X), then capping, then 20% piperidine; (ii) standard DIC coupling of HMPA (2X); (iii) 0.4 M Fmoc-Aa and 1.2 M CDI in DMF (4X), then, capping, then 20% piperidine; (iv) standard DIC coupling of Fmoc-Aa (2X), then capping, then 20% piperidine; (v) 0.5 M photocleavable linker, 0.5 M Et_3N in DMF (3X); (vi) 2% N_2H_4; (vii) 0.05 M BOP, 0.05 M 6-Cl-HOBt and 0.1 M DIEA in DMF (2X); (viii) modified Reagent K. (b) Screening and imaging strategy using CVIA (a substrate) and CVIL (nonsubstrate) as examples. Post-reaction colorization was accomplished by click reaction with biotin-azide followed by incubation with SP-AP. Colorization occurred upon the addition of BCIP.

4.3 Library screening using *R. norvegicus* PFTase

In previously reported work, Fierke and coworkers studied the specificity of

rPFTase by synthesizing and assaying 213 Ca_1a_2X-box containing peptide sequences found in the human proteome.[112] That sparse sampling of the total sequence space available from varying three residues (8000 possibilities) was quite useful for understanding prenylation specificity in the context of human biochemistry. However, such a study cannot reveal the complete specifity profile of the enzyme given the limited sampling. To complement that work, we chose to systematically assess the substrate specificity of mammalian PFTase at the a_2 and X positions by preparing and screening two peptide libraries (19×20 CVa_2X library and 19×20 CCa_2X library) using *R. norvegicus* PFTase (rPFTase). These libraries densely sample the possible range of sequence space from the last two positions. Inspection of the data provided in Figure 4.3 confirms that many of the sequences found in the human genome including CVIQ, CVIM, CVIS, CVTQ, CVVQ, CVVM and CVVS are efficiently prenylated as was previously reported by Fierke and coworkers and references therein.[112] However, a number of additional sequences not present in the human genome were prenylated with high efficiency including CVIA, CVIT, CVTS, CVVA, CVCS and others. This general observation is consistent with previous reports that rPFTase can prenylate peptide sequences not present in the genome and most of these have been demonstrated to be substrates using the *in vitro* fluorescence assay.[119,120] Some of these sequences are found in the genomes of other organisms including *S. pombe*

(CVIA), *S. commune* (CVVA), *D. menanogaster* (CVSS), *A. thaliana* (CVTA), and *M. musculus* (CVAS) whose prenylated proteins have not been studied in detail. Of potential greater significance are sequences found in human pathogens including CVFQ (*S. dysenteriae*) and CVTH (*B. circulans*) as well as various herpesviruses (CVSS and CVPF). Recent reports suggest that in some cases, bacterially and virally encoded proteins must be prenylated by endogenous human prenyltransferases for pathogenesis.[121-123] These results highlight the utility of this library approach as a tool for rapidly discovering/confirming new substrates.

a_2 / X	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V
E	2 3	2 3	3 3	2 3	2 4	1 2	2 2	1 3	1 4	2 7	1 3	2 3	2 3	2 4	1 4	2 5	2 3	3 2	2 3	2 6
Q	42 34	3 3	23 23	2 2	42 50	21 26	1 3	5 9	41 36	82 84	53 46	4 3	6 16	59 64	8 17	54 40	82 86	3 8	37 55	85 106
D	2 2	2 2	2 3	2 4	3 4	3 4	3 3	3 3	3 3	3 4	3 3	3 3	2 2	2 2	2 3	2 3	1 4	1 2	1 2	1 3
N	2 2	2 2	3 2	1 1	7 4	1 2	2 2	2 2	2 3	53 20	3 3	2 3	2 4	2 4	2 4	2 4	23 7	3 3	4 3	39 18
R	2 4	2 3	3 4	2 4	3 5	2 3	1 2	2 3	1 2	1 3	2 3	2 3	3 4	1 3	1 2	1 3	2 3	2 3	2 3	2 3
K	3 3	2 3	3 4	3 4	4 5	3 6	4 5	6 4	5 3	6 3	5 4	3 4	4 5	4 5	4 6	3 4	3 3	2 3	2 2	3 2
H	3 3	3 3	4 4	2 3	6 5	2 4	2 3	3 3	2 3	51 18	2 3	2 4	2 3	2 4	2 4	19 7	52 22	3 3	4 4	11 13
A	4 5	3 4	3 4	2 3	15 15	2 4	1 3	1 2	6 5	73 33	5 4	3 3	3 3	26 15	2 5	3 5	38 8	2 3	4 8	91 50
V	3 3	2 2	2 4	2 3	5 5	2 3	1 3	3 3	4 3	32 8	4 4	6 5	5 5	6 6	4 6	4 5	20 6	2 2	2 2	42 7
I	2 3	3 3	3 3	3 3	3 5	2 4	2 2	2 3	2 3	6 3	2 3	4 3	3 3	2 3	3 3	2 4	2 3	2 3	2 3	8 4
L	3 5	4 4	3 4	3 5	3 6	2 6	1 5	2 3	2 3	5 3	3 3	5 3	3 3	3 2	2 3	2 2	3 3	2 3	2 2	16 3
F	64 23	3 2	4 2	3 3	30 24	5 4	2 2	4 3	2 3	17 8	4 4	3 3	3 3	3 4	34 13	21 10	53 28	2 3	2 3	35 16
Y	2 5	2 3	3 6	2 3	6 14	2 5	1 3	2 3	2 6	21 12	1 3	2 2	3 3	3 4	3 2	4 8	18 15	1 3	2 6	17 15
W	3 4	3 4	2 4	2 3	3 4	2 4	2 3	2 3	2 3	3 2	2 3	3 4	3 4	3 3	2 3	2 2	2 2	2 2	1 2	2 2
G	2 2	2 2	2 4	2 2	6 3	2 2	2 2	2 3	2 4	11 13	3 4	3 4	3 3	3 3	3 3	3 4	3 5	2 3	2 2	11 6
C	4 3	3 3	3 3	3 3	12 8	2 2	2 2	3 3	3 3	36 16	11 3	3 3	4 3	11 3	8 3	12 4	43 12	2 3	2 4	39 30
M	15 7	2 3	4 3	2 3	24 21	4 4	2 3	2 3	8 8	65 67	9 6	3 3	2 4	21 21	3 2	6 4	51 19	2 2	12 11	85 56
S	47 3	2 2	11 2	2 3	74 31	51 4	2 2	6 3	14 3	90 67	47 10	3 3	4 3	22 4	26 5	38 5	90 27	3 3	3 3	90 77
T	4 3	3 3	3 4	2 4	12 4	2 2	2 2	3 2	4 3	66 10	3 2	2 1	2 2	11 3	2 2	4 2	41 3	2 2	2 3	61 19

Figure 4.3 Heat map representation of the extent of farnesylation of a RAGCVa_2X and RAGCCa_2X library of peptides by rPFTase. Columns represent the a_2 positions and rows represent the X positions. For each box, there are two numbers. The left and right values are spot intensities obtained with the CVa_2X and CCa_2X libraries, respectively. For comparison, the intensity of both libraries was normalized relative to that observed with the sequence CIIS. Each library was synthesized and screened at least two times and the average color intensities were color coded to produce the heat map. The intensities below 34% are shown in white,

those between 34 to 66% are shown in yellow, and those above 66% are shown in red. Diagonally shaded boxes represent sequences that are found in the human or rat genome.

Previously Fierke and coworkers reported that peptides containing cysteine as the a_1 position tend to have lower multiple turnover reactivity.[112] That inspired us to study CCa_2X peptides more thoroughly using a CCa_2X library. Since we had not previously prepared libraries containing adjacent Cys residues, a number of spots were selected and the peptides were cleaved from the membrane via UV irradiation and analyzed by MALDI MS. The spots gave the expected m/z values for the corresponding target peptides with an acceptable signal-to-noise ratio except when X was C, G or W. In screening the CCa_2X library, fewer positive spots were detected (compared to our observations with the CVa_2X library) indicating that generally CCa_2X peptides have lower reactivity. It was previously reported that the a_1 position of the Ca_1a_2X motif is more solvent exposed than the a_2 and X positions, and that there are no apparent restrictions at this position.[98] However, our screening results show that the a_1 position does have an effect on PFTase specificity, and PFTase discriminates against sequences with Cys as the a_1 residue unless the a_2 position is Ile or Val, and the X position is Gln, Met or Ser. *In vitro* enzymatic assay of selected peptides also confirmed these screening results (Tables 4.3). The observations described here are consistent with previously reported *in vivo* farnesylated protein

sequences. For example, it has been reported that the tyrosine phosphatases PRL-1, -2, -3 associate with the cellular membrane or endosome in a farnesylation-dependent manner.[124] Their C-terminal sequences are CCIQ, CCVQ and CCVM, respectively which also showed high reactivity in our screening analysis. As was noted above for the CVa_2X library, the CCa_2X library also contains numerous sequences that do not exist in the human genome but are still modified with high efficiency by rPFTase including CCIM, CCIS, CCTQ, and CCVS. The sequence CCSQ is particularly interesting since it occurs in the genome of the pathogenic bacteria *N. sennetsu* that resides within human cells; as noted above, prenylation of that protein by endogenous farnesyltransferase may be essential for pathogenesis. Finally, it should be noted that some proteins with C-terminal CCa_2X sequences are farnesylated or geranylgeranylated on the first Cys and then palmitoylated at the second Cys (the a_2 position) in lieu of proteolytic removal of the a_1a_2X tripeptide.[125] Thus, this study using a CCa_2X library paves the way for the use of such libraries to probe palmitoylation specificity. For this, a CCa_2X library could be prenylated chemically using orthogonally protected Cys[34,35] or enzymatically with PFTase or PGGTase-I and then screened for enzymatic palmitoylation using an alkyne-containing palmitoyl-CoA analogue.[126]

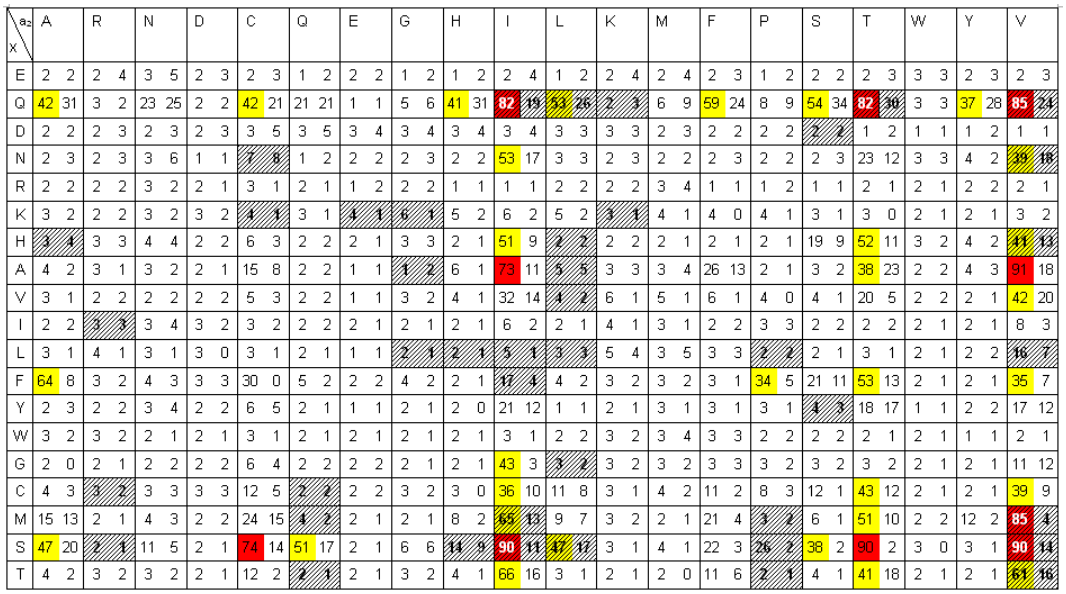
Overall, the results from screening CVa_2X and CCa_2X libraries confirm

many of the previously reported observations concerning rPFTase specificity. However, importantly, the use of these peptide libraries has allowed us to identify numerous sequences not present in the human genome that are efficient substrates for rPFTase. The discovery of such sequences may be useful for biotechnology applications in which Ca_1a_2X box sequences are appended onto proteins for site-specific labeling and builds upon results reported by the Fierke[127] and Hougland[128] groups on altering peptide specificity and work from Nguyen et al[97] on modulating isoprenoid substrate recognition. The ability of rPFTase to recognize sequences not found in the genome could serve as an important step in the development of orthogonal PFTase variants. Additionally given that many of the PFTase inhibitors created to date are peptidomimetic molecules based on Ca_1a_2X box sequences,[129,130] the insights obtained from the peptide libraries described here should be useful for inhibitor design. Finally, the discovery of substrate sequences present in the genomes of bacteria and viruses may shed light on their mechanisms of pathogenesis and suggest new approaches for therapeutic intervention.

4.4 Statistical analysis of library screening method and enzyme concentration dependence

In interpreting the data from the studies described above, we focused on the

medium (yellow) and high (red) intensity spots. To assess the validity of those measurements, statistical analysis was performed using the data from the CVa_2X library. Figure 4.4 shows the average spot intensity (from three separate experiments) at each position along with the standard deviation. Comparison of the percent standard deviation from the mean for each of the three categories (low, medium and high) gives values of 73, 35 and 18%, respectively. Further calculation for the 95% confidence interval for the same three categories yields values of 186, 88 and 44% suggesting that there is insufficient certainty (the uncertainty is larger than the values measured) in the data for the low intensity spots. Thus, this analysis suggests that spots that appear as yellow (medium intensity) or red (high intensity) are highly likely to be *bona fide* substrates. In contrast, it is not possible to differentiate with certainty whether spots that appear as white in the heat maps are poor substrates or not substrates at all.



X \ a_2	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V
E	2 2	2 4	3 5	2 3	2 3	1 2	2 2	1 2	1 2	2 4	1 2	2 4	2 4	2 3	1 2	2 2	2 3	3 3	2 3	2 3
Q	42 31	3 2	23 25	2 2	42 21	21 21	1 1	5 6	41 31	82 19	53 26	2 3	6 9	59 24	8 9	54 34	82 30	3 3	37 28	85 24
D	2 2	2 3	2 3	2 3	3 5	3 5	3 4	3 4	3 4	3 4	3 3	3 3	2 3	2 2	2 2	2 4	1 2	1 1	1 2	1 1
N	2 3	2 3	3 6	1 1	7 8	1 2	2 2	2 3	2 2	53 17	3 3	2 3	2 2	2 3	2 2	2 3	23 12	3 3	4 2	39 18
R	2 2	2 2	3 2	2 1	3 1	2 1	1 2	2 2	1 1	1 1	2 2	2 2	3 4	1 1	1 2	1 1	2 1	2 1	2 2	2 1
K	3 2	2 2	3 2	3 2	4 1	3 1	4 1	6 1	5 2	6 2	5 2	3 1	4 1	4 0	4 1	3 1	3 0	2 1	2 1	3 2
H	3 4	3 3	4 4	2 2	6 3	2 2	2 1	3 3	2 1	51 9	2 2	2 2	2 1	2 1	2 1	19 9	52 11	3 2	4 2	11 13
A	4 2	3 1	3 2	2 1	15 8	2 2	1 1	1 2	6 1	73 11	5 5	3 3	3 4	26 13	2 1	3 2	38 23	2 2	4 3	91 18
V	3 1	2 2	2 2	2 2	5 3	2 2	1 1	3 2	4 1	32 14	4 2	6 1	5 1	6 1	4 0	4 1	20 5	2 2	2 1	42 20
I	2 2	3 3	3 4	3 2	3 2	2 2	2 1	2 1	2 1	6 2	2 1	4 1	3 1	2 2	3 3	2 2	2 2	2 1	2 1	8 3
L	3 1	4 1	3 1	3 0	3 1	2 1	1 1	2 1	2 1	5 1	3 3	5 4	3 5	3 3	2 2	2 1	3 1	2 1	2 2	16 7
F	64 8	3 2	4 3	3 3	30 0	5 2	2 2	4 2	2 1	17 4	4 2	3 2	3 2	3 1	34 5	21 11	53 13	2 1	2 1	35 7
Y	2 3	2 2	3 4	2 2	6 5	2 1	1 1	2 1	2 0	21 12	1 1	2 1	3 1	3 1	3 1	4 3	18 17	1 1	2 2	17 12
W	3 2	3 2	2 1	2 1	3 1	2 1	2 1	2 1	2 1	3 1	2 2	3 2	3 4	3 3	2 2	2 2	2 1	2 1	1 1	2 1
G	2 0	2 1	2 2	2 2	6 4	2 2	2 2	2 1	2 1	43 3	3 4	3 2	3 2	3 3	3 2	3 2	3 2	2 1	2 1	11 12
C	4 3	3 2	3 3	3 3	12 5	2 2	2 2	3 2	3 0	36 10	11 8	3 1	4 2	11 2	8 3	12 1	43 12	2 1	2 1	39 9
M	15 13	2 1	4 3	2 2	24 15	4 2	2 1	2 1	8 2	65 13	9 7	3 2	2 1	21 4	3 2	6 1	51 10	2 2	12 2	85 4
S	47 20	2 1	11 5	2 1	74 14	51 17	2 1	6 6	14 9	90 11	47 17	3 1	4 1	22 3	26 2	38 2	90 2	3 0	3 1	90 14
T	4 2	3 2	3 2	2 1	12 2	4 1	2 1	3 2	4 1	66 16	3 1	2 1	2 0	11 6	2 1	4 1	41 18	2 1	2 1	64 16

Figure 4.4 Heat map representation of the extent of farnesylation of a RAGCVa_2X library by rPFTase. Columns represent specific a_2 positions and rows represent specific X positions. For each box, there are two numbers. The left and right values are spot intensities and standard deviations obtained from three screening experiments, respectively. For comparison, the intensities of both libraries were normalized relative to that observed with the sequence CIIS. The intensities below 34% are shown in white, those between 34 to 66% are shown in yellow, and those above 66% are shown in red. Diagonally shaded boxes indicate sequences that are found in the human or rat genome.

The statistical analysis described above suggests that it is not possible to conclude anything concerning the low intensity substrates. To examine this in greater detail, the effect of using a higher enzyme concentration was explored. Thus, the 380-member CCa_2X library was screened using an enzyme concentration 3-fold higher than that used for the earlier experiments. Figure 4.5 summarizes the results obtained from screening the CCa_2X library at the two different concentrations. Inspection of that data reveals several interesting observations. Most importantly, numerous spots that fell into the low intensity category at low enzyme concentration

moved into the medium or high category when the higher enzyme concentration was used including several whose intensity increased more than 2-fold. In some cases, sequences that were in the low intensity category increased significantly (more than 3-fold) but that increase was not sufficient to move them into the medium class given their low initial intensity. Of those, it should be noted that CCQS, CCIL, CCIF, CCGY, CCKT, CCPC and CCVL are present in the human genome. These results suggest that those sequences are real but poor substrates for the rPFTase. That conclusion is supported by data from Hougland et al[112] who did observe catalytic activity with the sequences CCIF and CCIL using *in vitro* fluorescence assays although it was 15- and 140-fold lower, respectively, compared with that for optimal substrates. In contrast, for most of the white spots, no significant increase in spot intensity was observed when the enzyme concentration was increased 3-fold, indicating that those sequences are either extremely poor substrates or not substrates at all in terms of multiple-turnover reactions. An alternative explanation is that some of these sequences may be single-turnover substrates under these *in vitro* conditions.[112] Efforts to explore the screening of these libraries using stoichiometric quantities of enzyme to reveal single-turnover substrates are in progress. Overall, this experiment demonstrates how the dynamic range of this screening method can be increased by raising the enzyme concentration and be used to detect activities that are more than

100-fold lower than that manifested by the fastest substrates.

a_2 / X	A		R		N		D		C		Q		E		G		H		I		L		K		M		F		P		S		T		W		Y		V	
E	3	4	3	6	3	6	3	4	4	3	2	3	2	3	3	5	4	5	7	6	3	4	3	5	3	6	4	5	4	4	5	4	3	6	2	7	3	7	6	6
Q	34	24	3	10	23	11	2	4	50	26	26	11	3	3	9	4	36	23	84	56	46	29	3	8	16	8	64	55	17	18	40	41	86	76	8	7	55	44	106	82
D	2	5	2	5	3	6	4	6	4	5	4	6	3	7	3	7	3	7	4	7	3	5	3	6	2	3	2	4	3	3	3	4	4	4	2	4	2	4	3	4
N	2	5	2	7	2	4	1	3	4	7	2	5	2	4	2	5	3	5	20	36	3	4	3	9	4	8	4	8	4	7	4	7	7	16	3	7	3	8	18	25
R	4	10	3	6	4	8	4	7	5	10	3	8	2	5	3	11	2	9	3	8	3	7	3	5	4	7	3	8	2	6	3	7	3	10	3	8	3	10	3	7
K	3	7	3	5	4	11	4	8	5	10	6	10	5	9	4	13	3	10	3	11	4	9	4	8	5	13	5	13	6	13	4	8	3	9	3	10	2	10	2	7
H	3	4	3	6	4	5	3	4	5	7	4	6	3	5	3	6	3	6	18	37	3	8	4	10	3	11	4	9	4	7	7	14	22	36	3	7	4	9	13	22
A	5	9	4	8	4	7	3	4	15	12	4	7	3	3	2	6	5	12	33	52	4	4	3	7	3	5	15	29	5	6	5	7	8	26	3	5	8	17	50	64
V	3	9	2	8	4	10	3	7	5	11	3	7	3	6	3	7	3	9	8	21	4	7	5	13	5	7	6	10	6	6	5	8	6	19	2	5	2	7	7	30
I	3	6	3	7	3	5	3	4	5	7	4	7	2	5	3	6	3	6	3	15	3	5	3	9	3	9	3	7	3	6	4	6	3	8	3	6	3	6	4	10
L	5	7	4	9	4	7	5	6	6	9	6	6	5	5	3	6	3	7	3	12	3	4	3	5	3	5	2	4	3	6	2	5	3	6	3	5	2	6	3	23
F	23	56	2	8	2	13	3	6	24	33	4	11	2	7	3	12	3	8	8	25	4	10	3	10	3	9	4	10	13	39	10	29	28	68	3	7	3	8	16	35
Y	5	8	3	7	6	12	3	4	14	18	5	13	3	5	3	11	6	24	12	59	3	13	2	7	3	18	4	25	2	12	8	24	15	39	3	6	6	23	15	36
W	4	7	4	8	4	6	3	6	4	6	4	6	3	5	3	6	3	9	2	7	3	4	4	6	4	5	3	7	3	8	3	4	2	4	2	6	2	5	2	6
G	2	6	2	8	2	8	2	8	3	9	2	8	2	6	3	8	4	8	13	33	4	6	4	9	3	7	3	8	3	6	4	5	5	11	3	7	2	4	6	23
C	3	6	3	7	3	5	3	3	8	22	2	5	2	4	3	5	3	8	16	37	3	7	3	6	3	9	3	20	3	10	4	18	12	39	3	7	4	9	30	34
M	7	20	3	5	3	11	3	4	21	50	4	13	3	5	3	8	8	21	67	87	6	12	3	6	4	7	21	48	2	7	4	10	19	56	2	6	11	50	56	102
S	3	28	2	6	2	10	3	4	31	54	4	28	2	5	3	9	3	15	67	81	10	25	3	8	3	8	4	23	5	41	5	25	27	77	3	6	3	10	77	87
T	3	6	3	6	4	7	4	5	4	12	2	6	2	4	2	5	3	8	10	18	2	6	1	7	2	7	3	16	2	5	2	8	3	32	2	6	3	7	19	48

Figure 4.5 Evaluation of the extent of farnesylation of a RAGCCa_2X library of peptides using different concentrations of rPFTase. Columns represent specific a_2 positions and rows represent specific X positions. For each box, there are two values. The left and right numbers indicate [rPFTase] = 55 ug/mL and 182 ug/mL results, respectively. Color intensities were quantified by Image J software. For comparison, the intensities in the CCa_2X library were normalized relative to that observed with the CIIS sequence. The library was synthesized and screened at least two times and the average color intensities were color-coded. The intensity below 34% is shown in white. The intensity between 34 to 66% is shown in yellow. The intensity above 66% is shown in red. Shaded boxes indicate the peptide sequences that are found in the human or rat genome.

4.5 Comparison with direct kinetic analysis and bioinformatics methods

Before using the libraries reported here in additional experiments, the results obtained with them in the experiments described above were compared with data obtained using two other methodologies. First the intensity data obtained from the CVa_2X library screening was compared with data obtained via direct kinetic analysis using individually synthesized peptides. Hougland et al[120] reported such data for a

series of peptides (n = 73) based on CVa_2X where X was restricted to A, M, Q and S. A plot of the library screening intensity versus k_{cat}/K_M gave a statistically significant correlation (Figure 4.6, $p < 0.001$) although there is significant deviation for some sequences. Comparison of data from individual sequences after binning the k_{cat}/K_M data into low, medium and high values (Figure 4.7) allows the correlation to be easily visualized. Inspection of that data reveals that methionine-containing sequences (with M at the a_2 or X position) correlate particularly poorly which may be due to oxidation to the sulfoxide during the synthesis and visualization process. In the MS analysis of selected methionine-containing peptide sequences, the sulfoxide is either the only detectable species or the major product present (see MS data for CCIM, Figure S4.1); similar results were noted in our earlier library work with yeast PFTase (with the sequence CVIM)[39] and in other libraries currently being prepared including the sequences CKIM, CDIM, CTIM, CFIM, CGIM (data not shown). The library screening data obtained here was also compared with predictions made using PrePS, a web-based bioinformatics program that scores sequences for their potential to be PFTase substrates.[131] A plot of the library screening intensity versus PrePS score gave a statistically significant correlation (Figure 4.8, $p < 0.005$). Comparison of data from individual sequences after binning the PrePS scores into low, medium and high values (Figure 4.9) shows that many sequences observed as positives in the library screening

are predicted to be PFTase substrates by PrePS. As noted for the k_{cat}/K_M data, methionine-containing sequences also correlate poorly. Conversely, the library screening reveals a number of sequences that are enzyme substrates that are not predicted by PrePS including CVFQ, CVSQ and CVYQ. Those sequences are particularly noteworthy since they were also identified as substrates in assays with individual peptides by Fierke and coworkers.[120] In general, the comparisons noted above show that the data obtained with the peptide libraries described herein reproduces many of the features of substrate recognition previously reported for PFTase and thus underscores the validity of using these libraries for rapidly assessing aspects of PFTase substrate specificity.

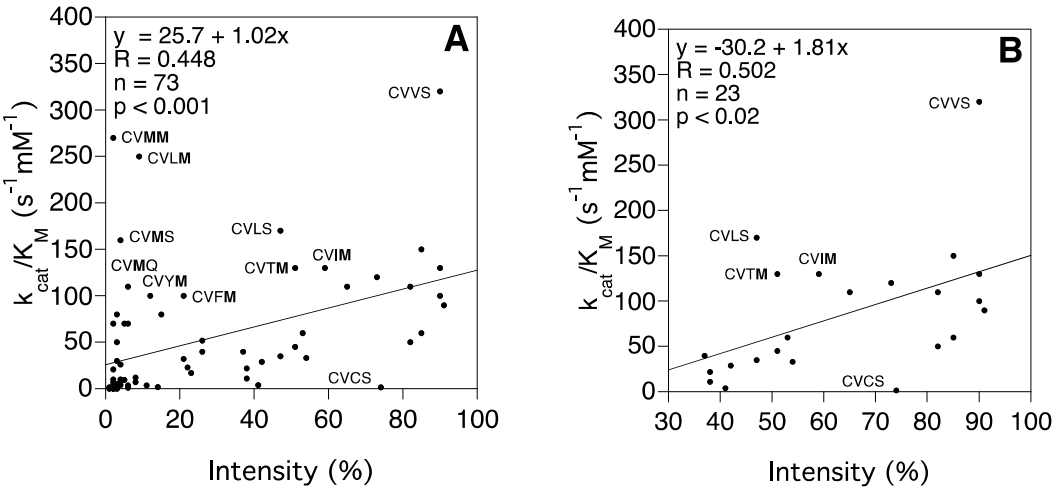


Figure 4.6 Comparison of intensity values obtained via library screening with k_{cat}/K_M values obtained from kinetic analysis performed with individual synthetic peptides reported by Fierke and coworkers. (A) Comparison using all data reported by Fierke and coworkers. (B) Comparison limited to library screening data that yielded intensity values falling into the

medium (34-66%) and high (67-100%) intensity categories. Data from certain methionine-containing sequences and other outliers are indicated. Kinetic data obtained from Hougland et al.[120]

X \ a_2	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V
Q	42 29	3 N	23 17	2 <1	42 N	21 32	1 <1	5 9.4	41 4	82 110	53 60	2 1.9	6 110	59 130	8 12	54 33	82 50	3 50	37 40	85 60
A	4 3.9	3 N	3 <1	2 <1	15 N	2 6.8	1 <1	1 1.6	6 1.1	73 120	5 70	3 <1	3 80	26 40	2 3.4	3 3.2	38 11	2 10	4 7	91 90
M	15 80	2 N	4 26	2 <1	24 N	4 10	2 <1	2 21	8 7	65 110	9 250	3 1.3	2 270	21 100	3 30	6 70	51 130	2 70	12 100	85 150
S	47 35	2 N	11 3.6	2 <1	74 1.4	51 45	2 <1	6 3.5	14 1.9	90 130	47 170	3 <1	4 160	22 23	26 52	38 22	90 100	3 5	3 2.1	90 320

Figure 4.7 Comparison of screening intensities obtained using the RAGCVa_2X library of peptides with k_{cat}/K_M values ($mM^{-1}s^{-1}$) from Hougland et al.[120] For each box, two numbers are listed. The left and right values are spot intensities and k_{cat}/K_M values, respectively. The intensities below 34% are shown in white. The intensities between 34 to 66% are shown in yellow. The intensities above 66% are shown in red. The k_{cat}/K_M values below 19 are shown in white. Those between 20 to 100 are shown in yellow and those above 100 are shown in red.

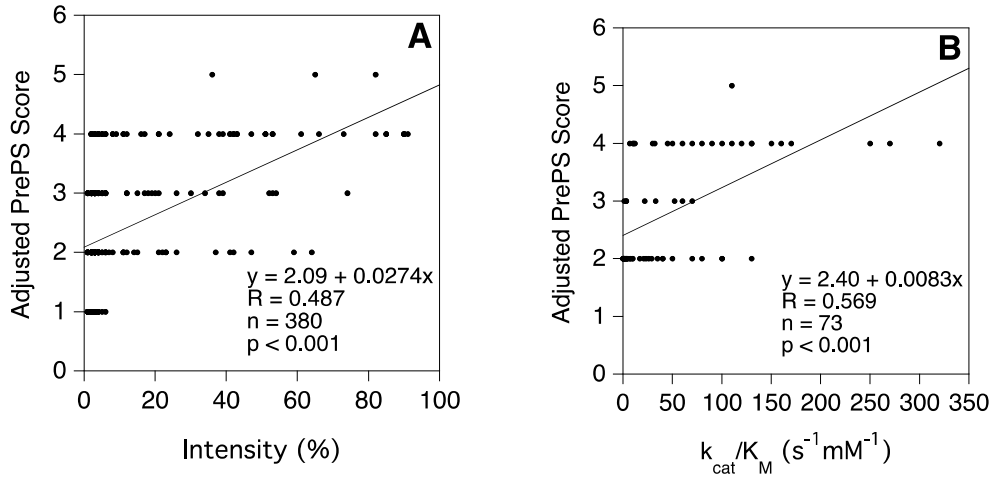


Figure 4.8 Comparison of PrePS scores with intensity values obtained via library screening or k_{cat}/K_M values obtained from kinetic analysis performed with individual synthetic peptides reported by Fierke and coworkers.[120] (A) Comparison of PrePS scores with intensity values obtained via library screening reported here with the CVa_2X library. (B) Comparison of PrePS scores with k_{cat}/K_M values reported by Fierke and coworkers.[120] The kinetic data used for (B) was obtained from Hougland et al.[3] To convert PrePS scores containing (+) and (-), the following scale was created: (+++) = 5; (++) = 4; (+) = 3; (-) = 2; (- -) = 1; (- - -) = 0. To obtain PrePS scores, the H-ras sequence KLNPPDESGPGCMRAGCVa_2X was used where a_2X are the residues at the a_2 and X positions, respectively, since the program requires a

minimum peptide length of 20 residues.

(a)

a_2 / X	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V
E	2 -2	2 -2	3 -2	2 -2	2 -1	1 +1	2 -2	1 -2	1 -2	2 +1	1 +1	2 -2	2 -1	2 -2	1 -1	2 -1	2 +1	3 -2	2 -2	2 +1
Q	42 -1	3 -1	23 -1	2 -1	42 +2	21 +2	1 -1	5 -1	41 -1	82 +3	53 +2	2 -1	6 +2	59 -1	8 +2	54 +1	82 +2	3 -1	37 -1	85 +2
D	2 -2	2 -2	2 -2	2 -2	3 -1	3 -1	3 -2	3 -2	3 -2	3 +1	3 -1	3 -2	2 -1	2 -2	2 -1	2 -1	1 -1	1 -2	1 -2	1 +1
N	2 -2	2 -2	3 -2	1 -2	1 -1	1 +1	2 -2	2 -2	2 -2	53 +1	3 +1	2 -2	2 -1	2 -2	2 -1	2 -1	23 -1	3 -2	4 -2	39 +1
R	2 -2	2 -2	3 -2	2 -2	3 -1	2 +1	1 -2	2 -2	1 -2	1 +1	2 +1	2 -2	3 -1	1 -2	1 -1	1 -1	2 +1	2 -2	2 -2	2 +1
K	3 -1	2 -2	3 -2	3 -2	4 -1	3 +1	4 -2	6 -2	5 -2	6 +1	5 +1	3 -2	4 -1	4 -2	4 -1	3 -1	3 +1	2 -2	2 -2	3 +1
H	3 -1	3 -1	4 -1	2 -1	6 +1	2 +1	2 -2	3 -1	2 -1	51 +2	2 +2	2 -1	2 +1	2 -1	2 +1	19 +1	52 +1	3 -2	4 -1	41 +2
A	4 -1	3 -1	3 -1	2 -1	15 +1	2 +2	1 -1	1 -1	6 -1	73 +2	5 +2	3 -1	3 +2	26 -1	2 +1	3 +1	38 +2	2 -1	4 -1	91 +2
V	3 -1	2 -1	2 -1	2 -1	5 +1	2 +2	1 -1	3 -1	4 -1	32 +2	4 +2	6 -1	5 +1	6 -1	4 +1	4 +1	20 +1	2 -1	2 -1	42 +2
I	2 -1	3 -1	3 -1	3 -1	3 +1	2 +2	2 -1	2 -1	2 -1	6 +2	2 +2	4 -1	3 +1	2 -1	3 +1	2 +1	2 +2	2 -1	2 -1	8 +2
L	3 -1	4 -1	3 -1	3 -1	3 +1	2 +2	1 -1	2 -1	2 -1	5 +2	3 +2	5 -1	3 +1	3 -1	2 +1	2 +1	3 +2	2 -1	2 -1	16 +2
F	64 -1	3 -1	4 -1	3 -1	30 +1	5 +2	2 -1	4 -1	2 -1	17 +2	4 +2	3 -1	3 +1	3 -1	34 +1	21 +1	53 +2	2 -2	2 -1	35 +2
Y	2 -1	2 -2	3 -1	2 -2	6 +1	2 +1	1 -2	2 -1	2 -1	21 +2	1 +1	2 -2	3 +1	3 -2	3 +1	4 +1	18 +1	1 -2	2 -2	17 +1
W	3 -2	3 -2	2 -2	2 -2	3 -1	2 -1	2 -2	2 -1	2 -2	3 +1	2 -1	3 -2	3 -1	3 -2	2 -1	2 -1	2 +1	2 -2	1 -2	2 +1
G	2 -1	2 -1	2 -1	2 -2	6 +1	2 +1	2 -2	2 -1	2 -1	43 +2	3 +1	3 -1	3 +1	3 -1	3 +1	3 +1	3 +1	2 -1	2 -1	11 +2
C	4 -1	3 -1	3 -1	3 -1	12 +2	2 +2	2 -1	3 -1	3 -1	36 +3	11 +2	3 -1	4 +2	11 -1	8 +2	12 +1	43 +2	2 -1	2 -1	39 +2
M	15 -1	2 -1	4 -1	2 -1	24 +2	4 +2	2 -1	2 -1	8 -1	65 +3	9 +2	3 -1	2 +2	21 -1	3 +2	6 +1	51 +2	2 -1	12 -1	85 +2
S	47 -1	2 -1	11 -1	2 -1	74 +1	51 +2	2 -1	6 -1	14 -1	90 +2	47 +2	3 -1	4 +2	22 -1	26 +1	38 +1	90 +2	3 -1	3 -1	90 +2
T	4 -1	3 -1	3 -1	2 -1	12 +1	2 +2	2 -1	3 -1	4 -1	66 +2	3 +2	2 -1	2 +1	11 -1	2 +1	4 +1	41 +2	2 -1	2 -1	61 +2

(b)

a_2 / X	K	D	E	N	Q	S	G	H	T	A	P	Y	V	M	C	L	F	I	W
Q	1.9 -1	<1 -1	<1 -1	17 -1	32 +2	33 +1	9.4 -1	4 -1	50 +2	29 -1	12 +2	40 -1	60 +2	110 +2	nd +2	60 +2	130 -1	110 +3	50 -1
A	<1 -1	<1 -1	<1 -1	<1 -1	6.8 +2	3.2 +1	1.6 -1	1.1 -1	11 +2	3.9 -1	3.4 +1	7 -1	90 +2	80 +2	nd +2	70 +2	40 -1	120 +2	10 -1
M	1.3 -1	<1 -1	<1 -1	26 -1	10 +2	70 +1	21 -1	7 -1	130 +2	80 -1	30 +2	100 -1	150 +2	270 +2	nd +2	250 +2	100 -1	110 +3	70 -1
S	<1 -1	<1 -1	<1 -1	3.6 -1	45 +2	22 +1	3.5 -1	1.9 -1	100 +2	35 -1	52 +1	2.1 -1	320 +2	160 +2	1.4 +2	170 +2	23 -1	130 +2	5 -1

Figure 4.9 (a) Comparison of screening intensities from a RAGCVa_2X library of peptides (left values) with PrePS predictions (right values). (b) Comparison of k_{cat}/K_M ($mM^{-1}s^{-1}$) values from Hougland paper[120] (left values) with PrePS predictions (right values). To obtain PrePS scores, the H-ras sequence KLNPPDESGPGCMRAGCVa_2X was used where a_2X are the residues at the a_2 and X positions, respectively. The intensities below 34% are shown white. Those between 34 to 66% are shown in yellow and those above 66% are shown in red. The k_{cat}/K_M values below 20 are shown in white. Those between 20 to 100 are shown in yellow and those above 100 are shown in red. For PrePS predictions, values of -2 to +1 are shown in white, values of +2 are shown in yellow and values of +3 are shown in red.

4.6 Examining the interplay between peptide specificity and isoprenoid length

Inspection of the crystal structure of the ternary complex of PFTase bound to a peptide substrate and isoprenoid analog reveals extensive contacts between the isoprenoid and the peptide.[98] Previously, Gibbs and coworkers observed in studies using a series of modified isoprenoid diphosphates that peptide substrate specificity varied depending on which particular isoprenoid was used.[132] That is, while some isoprenoids worked well for certain peptide sequences, other isoprenoids worked best with different peptide substrates. Since a number of isoprenoid analogues are currently being used to probe protein prenylation in living cells[99,101-103,133-135] and it would be useful to know whether modifications in the isoprenoid structure show systematic differences in peptide specificity. Similarly, isoprenoid analogues are being developed to disrupt interactions with prenylated proteins.[96] In those cases it will be important to know whether such analogues are incorporated into prenylated proteins with a specificity profile similar to farnesyl groups. While *in vitro* assays with individual peptides do allow the catalytic efficiencies of different isoprenoid substrates to be compared, such experiments do not allow global variations in peptide substrate efficiency to be studied. To address this question for alkyne probes currently being used for proteomic applications, we screened a 380 member CVa_2X library using three different isoprenoid analogues of varying length (C5-Alk, C10-Alk, C15-Alk, Figure 4.1). The screening data is summarized in Figure 4.10. From the

results, it can be seen that the specificity profiles obtained using the C10-Alk and C15-Alk substrates are similar. For example, CVCS, CVHQ, CVIQ, CVIM, CVIS, CVTQ, CVTS, CVVQ, CVVM and other sequences manifest comparable reactivity with C10-Alk and C15-Alk. In some cases, C10-Alk reacts preferentially compared with C15-Alk (CVCQ, CVMQ, CVLM and CVLS) while in other cases C15-Alk reacts preferentially compared with C10-Alk (CVAF, CVIA, CVIG and CVTF). In most cases, the observed differences are on the order of 2-fold making them relatively minor. However, in a few cases, more significant differences exist. Those include cases where C15-Alk exhibits preferential reactivity (CVAF, CVTF and CVVV) as well as a few instances where C10-Alk displays greater reactivity (CVQQ and CVMQ). In aggregate, similar results were obtained with these two different isoprenoids suggesting that probe choice does not skew results in proteomic investigations although care should be taken in quantitative proteomic experiments since significant variations in labeling do occur in a handful of cases.

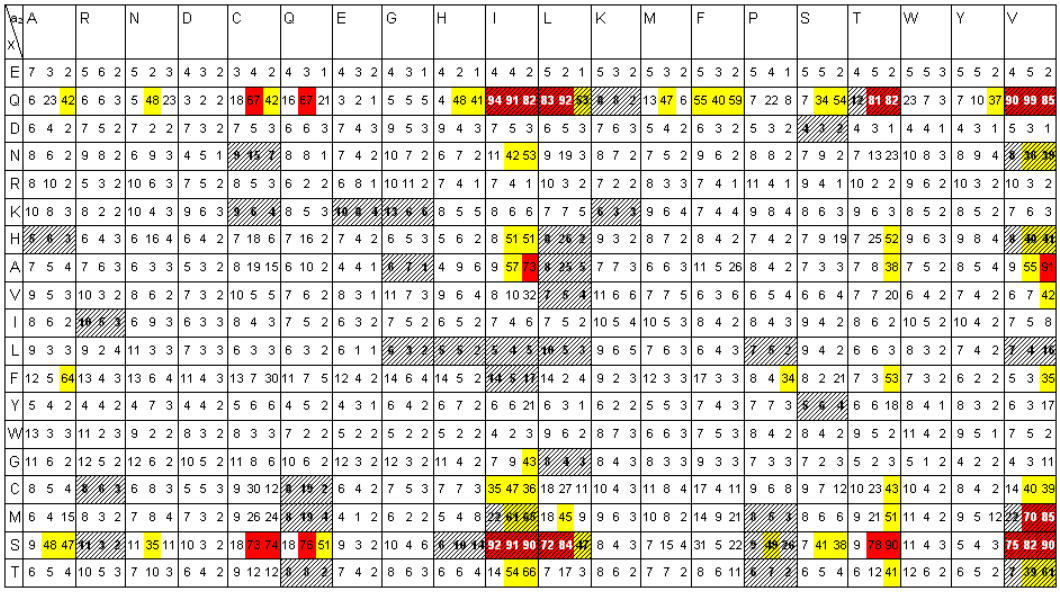


Figure 4.10 Heat map representation of the extent of farnesylation of a RAGCVa$_2$X library of peptides with different isoprenoid substrates. For each sequence, the left value shows the results using C5-Alk, the middle value using C10-Alk, the right value using C15-Alk. For comparison, the intensity was normalized relative to that observed with CVIS. Color coding and shading was performed as described for Figure 4.2.

In contrast to the above observations, experiments with the C5-Alk gave different results. First, it is clear that the C5-Alk is not an efficient substrate for PFTase since fewer positive spots were observed in the screening experiment. A comparison of the kinetic constants measured with an *in vitro* enzyme assay using FPP, C10-Alk and C15-Alk (Table 4.1) shows that while the catalytic efficiencies of C10-Alk and C15-Alk are comparable but lower than FPP. C5-Alk is a much poorer substrate with K_M 78-fold higher and k_{cat} 400-fold lower compared with FPP (Table 4.2); attenuated catalytic activity was previously noted with that analogue using yPFTase.[136] More importantly, the specificity profile with C5-Alk was significantly different from that

of the other probes. For example, CVIQ, CVIS, CVVQ and CVVS all showed high activity regardless of which isoprenoid probe was used. In contrast, CVCQ, CVHQ, CVTQ, CVAS, CVCS, and CVTS were highly reactive when using C10-Alk and C15-Alk but not in the presence of C5-Alk. Those results demonstrate that the peptide substrate selectivity obtained using C5-Alk differs significantly from that observed with C10-Alk and C15-Alk and suggests that this shorter probe is not suitable for proteomic analysis of farnesylated proteins since its peptide specificity profile is likely to be different than that for FPP. However, the different pattern of reactivity observed with C5-Alk may be useful for designing orthogonal prenyltransferases that recognize novel sequences. Overall, these experiments illustrate how C-terminal peptide libraries can be used to rapidly examine the effect of changes in isoprenoid structure on the prenyltransferase specificity.

isoprenoid	k_{cat} (s^{-1})	K_M (uM)[a]	k_{cat}/K_M ($uM^{-1}s^{-1}$)[a]	k_{cat}/K_M (rel.)
FPP	0.019 [4]	0.107 [4]	0.18	
(Literature value)	0.29 ± 0.01 [5]	1.5 ± 0.2 [5]	0.19	
FPP	0.21 ± 0.01	0.11 ± 0.03	1.9	1
C10-Alk	0.24 ± 0.01	0.61 ± 0.15	0.39	0.21
C15-Alk	0.08 ± 0.004	0.14 ± 0.06	0.57	0.31

[a]K_M in this case is the K_M for the isoprenoid substrate.

Table 4.1 Kinetic Constants for the rPFTase-Catalyzed Reaction of FPP and analogues with DsGCVLS peptides. Assays were performed with 2.0 μM DsGCVLS, 20 nM rPFTase, varying concentrations (0.25-16 μM) of FPP analogues, 50 mM Tris, pH 7.5, 5 mM DTT, 5

mM $MgCl_2$, 50 μM $ZnCl_2$, and 0.040 % (*w/v*) *n*-dodecyl-*β*-D-maltoside.

rPFTase	k_{cat} (s^{-1})	K_M (uM)[a]	k_{cat}/K_M ($s^{-1}uM^{-1}$)[a]
C5-Alk	$6.3(\pm0.4)\times10^{-5}$	12 ± 3	5.0×10^{-6}
FPP	$1.7(\pm0.2)\times10^{-2}$	2.5 ± 0.7	6.0×10^{-3}

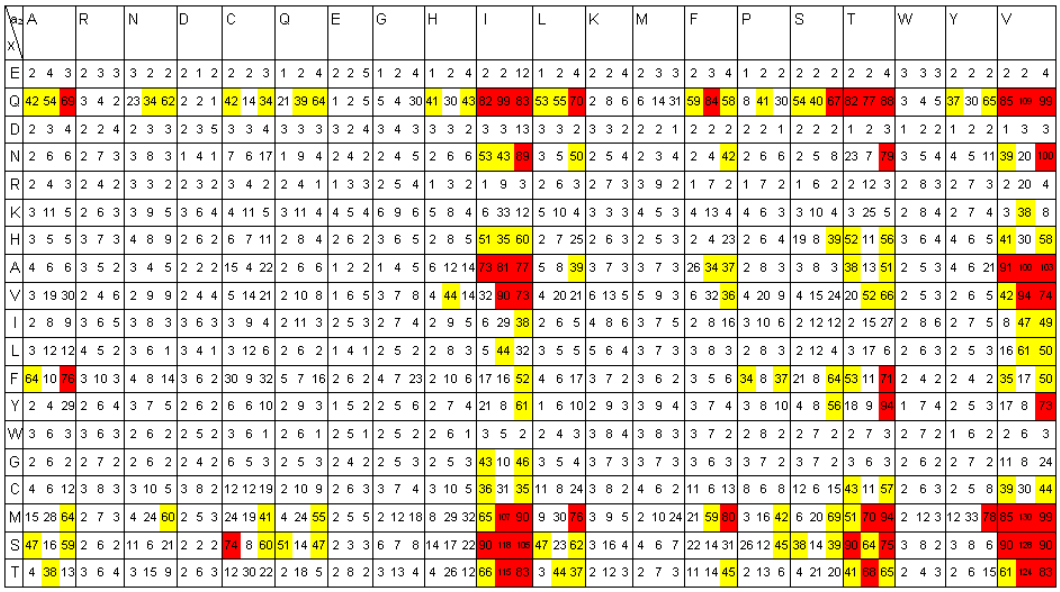
[a]K_M in this case is the K_M for the isoprenoid substrate.

Table 4.2 Evaluation of OrG-RTRCVIS peptides for *R. norvegicus* PFTase. Assays were performed at varying concentration of isoprenoid (0.2-20 uM FPP or 5-200 uM C5-Alk), 2.1 μM OrG-RTRCVIS peptide, 50 nM PFTase (for FPP) or 500 nM PFTase (for C5-Alk), 50 mM Tris, pH 7.5, 5 mM DTT, 5 mM $MgCl_2$, 50 μM $ZnCl_2$. Work reported from Poulter and coworkers indicates that OrG-GCVIA is a poorer substrate compared with DsGCVIA.[6,7]

4.7 Probing the specificity of PFTases from different species

Given the efficiency of the method described here for systematically and rapidly probing a large number of peptide sequences, we envisioned that it would be an effective strategy for examining the specificity of PFTase enzymes from different species. Thus C-terminal peptide libraries were used to compare the specificity between PFTases from three different species (*R. norvegicus, C. albicans* and *S. cerevisiae*). Accordingly, SPOT synthesis was employed to prepare the same 19×20 CVa_2X library described above followed by screening and visualization using C15-Alk as a substrate. The screening results are summarized in Figure 4.11. A random 60-member library with a_1 residues different from valine was also synthesized, screened, and visualized as noted above (see Figure 4.12). For these experiments, screening was performed using an equivalent amount of enzymatic activity by

normalizing for differences in reaction rate using DsGCVIA (2 uM) and saturating concentrations of C15-Alk (100 uM).



a2 \ X	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V
E	2 4 3	2 3 3	3 2 2	2 1 2	2 2 3	1 2 4	2 2 5	1 2 4	1 2 4	2 2 12	1 2 4	2 2 4	2 3 3	2 3 4	1 2 2	2 2 2	2 2 4	3 3 3	2 2 2	2 2 4
Q	42 54 69	3 4 2	23 34 62	2 2 1	42 14 34	21 39 64	1 2 5	5 4 30	41 30 43	82 99 83	53 55 70	2 8 6	6 14 31	59 84 58	8 41 30	54 40 67	82 77 88	3 4 5	37 30 65	85 109 99
D	2 3 4	2 2 4	2 3 3	2 3 5	3 3 4	3 3 3	3 2 4	3 4 3	3 3 2	3 3 13	3 3 2	3 3 2	2 2 1	2 2 2	2 2 1	2 2 2	1 2 3	1 2 2	1 2 2	1 3 3
N	2 6 6	2 7 3	3 8 3	1 4 1	7 6 17	1 9 4	2 4 2	2 4 5	2 6 6	53 43 88	3 5 50	2 5 4	2 3 4	2 4 42	2 6 6	2 5 8	23 7 79	3 5 4	4 5 11	39 20 100
R	2 4 3	2 4 2	3 3 2	2 3 2	3 4 2	2 4 1	1 3 3	2 5 4	1 3 2	1 9 3	2 6 3	2 7 3	3 9 2	1 7 2	1 7 2	1 6 2	2 12 3	2 8 3	2 7 3	2 20 4
K	3 11 5	2 6 3	3 9 5	3 6 4	4 11 5	3 11 4	4 5 4	6 9 6	5 8 4	6 33 12	5 10 4	3 3 3	4 5 3	4 13 4	4 6 3	3 10 4	3 25 5	2 8 4	2 7 4	3 38 8
H	3 5 5	3 7 3	4 8 9	2 6 2	6 7 11	2 8 4	2 6 2	3 6 5	2 8 5	51 35 60	2 7 25	2 6 3	2 5 3	2 4 23	2 6 4	19 8 39	52 11 56	3 6 4	4 6 5	41 30 58
A	4 6 6	3 5 2	3 4 5	2 2 2	15 4 22	2 6 6	1 2 2	1 4 5	6 12 14	73 81 77	5 8 39	3 7 3	3 7 3	26 34 37	2 8 3	3 8 3	38 13 51	2 5 3	4 6 21	91 100 103
V	3 19 30	2 4 6	2 9 9	2 4 4	5 14 21	2 10 8	1 6 5	3 7 8	4 44 14	32 90 73	4 20 21	6 13 5	5 9 3	6 32 36	4 20 9	4 15 24	20 52 66	2 5 3	2 6 5	42 94 74
I	2 8 9	3 6 5	3 8 3	3 6 3	3 9 4	2 11 3	2 5 3	2 7 4	2 9 5	6 29 38	2 6 5	4 8 6	3 7 5	2 8 16	3 10 6	2 12 12	2 15 27	2 8 6	2 7 5	8 47 49
L	3 12 12	4 5 2	3 6 1	3 4 1	3 12 6	2 6 2	1 4 1	2 5 2	2 8 3	5 44 32	3 5 5	5 6 4	3 7 3	3 8 3	2 8 3	2 12 4	3 17 6	2 6 3	2 5 3	16 61 50
F	64 10 76	3 10 3	4 8 14	3 6 2	30 9 32	5 7 16	2 6 2	4 7 23	2 10 6	17 16 52	4 6 17	3 7 2	3 6 2	3 5 6	34 8 37	21 8 64	53 11 71	2 4 2	2 4 2	35 17 50
Y	2 4 29	2 6 4	3 7 5	2 6 2	6 6 10	2 9 3	1 5 2	2 5 6	2 7 4	21 8 61	1 6 10	2 9 3	3 9 4	3 7 4	3 8 10	4 8 56	18 9 94	1 7 4	2 5 3	17 8 73
W	3 6 3	3 6 3	2 6 2	2 5 2	3 6 1	2 6 1	2 5 1	2 5 2	2 6 1	3 5 2	2 4 3	3 8 4	3 8 3	3 7 2	2 8 2	2 7 2	2 7 3	2 7 2	1 6 2	2 6 3
G	2 6 2	2 7 2	2 6 2	2 4 2	6 5 3	2 5 3	2 4 2	2 5 3	2 5 3	43 10 46	3 5 4	3 7 3	3 7 3	3 6 3	3 7 2	3 7 2	3 6 3	2 6 2	2 7 2	11 8 24
C	4 6 12	3 8 3	3 10 5	3 8 2	12 12 19	2 10 9	2 6 3	3 7 4	3 10 5	36 31 35	11 8 24	3 8 2	4 6 2	11 6 13	8 6 8	12 6 15	43 11 57	2 6 3	2 5 8	39 30 44
M	15 28 64	2 7 3	4 24 60	2 5 3	24 19 41	4 24 55	2 5 5	2 12 18	8 29 32	65 107 90	9 30 76	3 9 5	2 10 24	21 59 80	3 16 42	6 20 69	51 70 94	2 12 3	12 33 78	85 130 99
S	47 16 59	2 6 2	11 6 21	2 2 2	74 8 60	51 14 47	2 3 3	6 7 8	14 17 22	90 118 105	47 23 62	3 16 4	4 6 7	22 14 31	26 12 45	38 14 39	90 64 75	3 8 2	3 8 6	90 128 90
T	4 38 13	3 6 4	3 15 9	2 6 3	12 30 22	2 18 5	2 8 2	3 13 4	4 26 12	66 115 83	3 44 37	2 12 3	2 7 3	11 14 45	2 13 6	4 21 20	41 68 65	2 4 3	2 6 15	61 124 83

Figure 4.11 Heat map representation of the extent of farnesylation of a RAGCVa$_2$X library of peptides by three different PFTases. Rows represent a$_2$ position and columns represent X position. For each sequence (each box), there are three sections. The left, middle and right values indicated results obtained with *R. norvegicus*, *C. albicans, S. cerevisiae* PFTases, respectively. For comparison, the intensity was normalized relative to that observed with CVIS. Color coding was performed as described for Figure 4.2.

CTRK 4 4 2	CQRK 3 4 3	CKCI 6 21 3	CSEI 2 6 2	CIII 3 15 11	CKYI 4 12 3	CASL 2 13 3	CTIL 21 84 43	CSGL 2 13 3	CAIL 5 32 13
CIIL 2 28 11	CIDL 1 6 2	CPFW 1 6 2	CHDE 2 4 2	CRME 1 6 1	CKGE 1 5 1	CCLD 1 3 1	CHHD 2 4 2	CDDD 2 3 1	CALD 1 2 1
CWKD 2 3 1	CDPN 1 2 0	CRNR 1 3 1	CQGK 4 9 4	CLAK 3 10 3	CRVK 2 11 2	CIGK 3 10 4	CIHH 2 6 4	CYNA 1 4 2	CTVA 98 88 92
CSNA 3 5 3	CPKA 5 10 4	CYDA 3 6 3	CNDV 2 5 3	CMYV 5 12 4	CFIF 2 5 7	CIQF 5 6 7	CNAG 2 6 3	CCVC 17 9 6	CSIM 77 97 93
CCIM 44 42 42	CTIM 94 116 97	CIIM 82 90 83	CAIM 88 116 100	CTLM 26 58 97	CCPS 14 4 3	CDFS 2 3 2	CIIS 100 100 100	CCCS 30 4 9	CKQS 15 24 9
CIKS 4 10 4	CTDS 2 1 1	CKCT 11 47 5	CKQQ 38 77 46	CCIQ 95 62 57	CASQ 75 18 74	CDDY 1 3 1	CAPY 2 3 2	CADY 1 2 1	CPNY 2 3 1

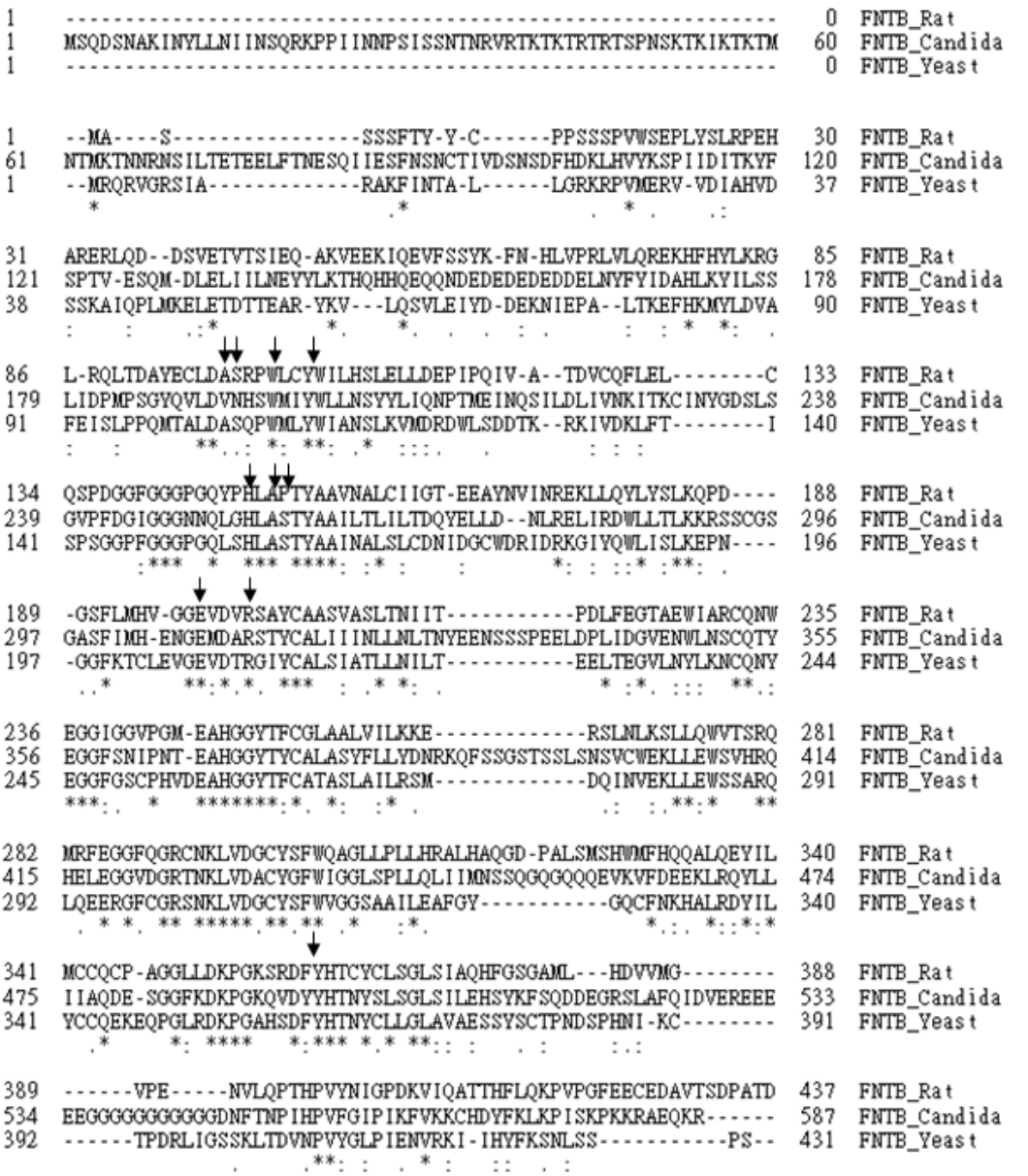
Figure 4.12 Evaluation of the extent of farnesylation of a random library of peptides by PFTases from three different species. For each box, there are three values. The left, middle and right values were obtained using *R. norvegicus*, *C. albicans, S. cerevisiae* PFTase, respectively. Color intensities were quantified by Image J software. For comparison, the intensities were normalized relative to that observed with the CIIS sequence. Color coding was performed as described for Figure 4.2.

In general, it can be seen that the three enzymes have similar peptide specificities (Figure 4.11). Thus, they all display a preference for Ile and Val at the a_2 position, and Gln, Ala, Met, Ser and Thr at the X position. This is not surprising, given their sequence similarity. BLAST analysis of the PFTase β-subunit shows that the residues contacting the peptide substrate are mostly conserved (Figure 4.13). From the numbers of active peptides recognized by the three enzymes (Figure 4.14), several conclusions can be made. First, yPFTase prenylates more sequences (84, 19%) than the other two enzymes making it the most promiscuous of the three. In contrast, the

CaPFTase is the most selective, efficiently prenylating only 47 (11%) out of the 440 examined; rPFTase exhibits similar overall specificity, prenylating 51 of the sequences (12%). Next, despite those differences in selectivity, rPFTase shares more substrates with yPFTase (51) than with CaPFTase (41). This observation is consistent with sequence alignment of the different genes. The β-subunit of rPFTase has higher similarity with yPFTase than with CaPFTase (58% vs 47%). Finally, there are significant differences between how the three different enzymes process identical substrates. Numerous examples occur where the *C. albicans* enzyme manifests higher relative reactivity with a given sequence. Those include CVHV, CVIV, CVLT, CVFM, CVPQ, CVTV, CVVV, CVVI, CVVL and CVVT which all show at least a 2-fold higher intensity with CaPFTase relative to rPFTase (see Figure 4.15a for data presented in graphical form). As an example, *in vitro* analysis using the peptide Ds-GCVVV reveals that CaPFTase exhibits a 6.2-fold greater k_{cat}/K_M value when compared with rPFTase (Table 4.3). Similar evidence for greater preference for yPFTase compared with rPFTase was also observed in this data. The sequences CVAM, CVNM, CVIV, CVLN, CVLA, CVLM, CVLT, CVFM, CVFT, CVPM, CVSF, CVSM, CVTN, CVTV, CVTY, CVYM, CVVN, CVVI, CVVL and CVVY all show at least a 2-fold higher intensity with yPFTase relative to rPFTase (see Figure 4.15b for data presented in graphical form). *In vitro* analysis using the peptide Ds-GCVTN

reveals that yPFTase exhibits a 11-fold greater k_{cat}/K_M value when compared with rPFTase (Table 4.3). While these differences reflect deviations in catalytic activity, they suggest that significant variations in sequence specificity occur between these enzymes. Such differences may be useful for the design of PFTase inhibitors that are selective for different species. Such inhibitors could be useful as drugs for the treatment of a number of diseases beyond cancer including malaria, leshmania and hepatitis.

(a)


FNTB_Rat
FNTB_Candida
FNTB_Yeast
Guide tree
FNTB_Candida
FNTB_Rat
FNTB_Yeast

(b)

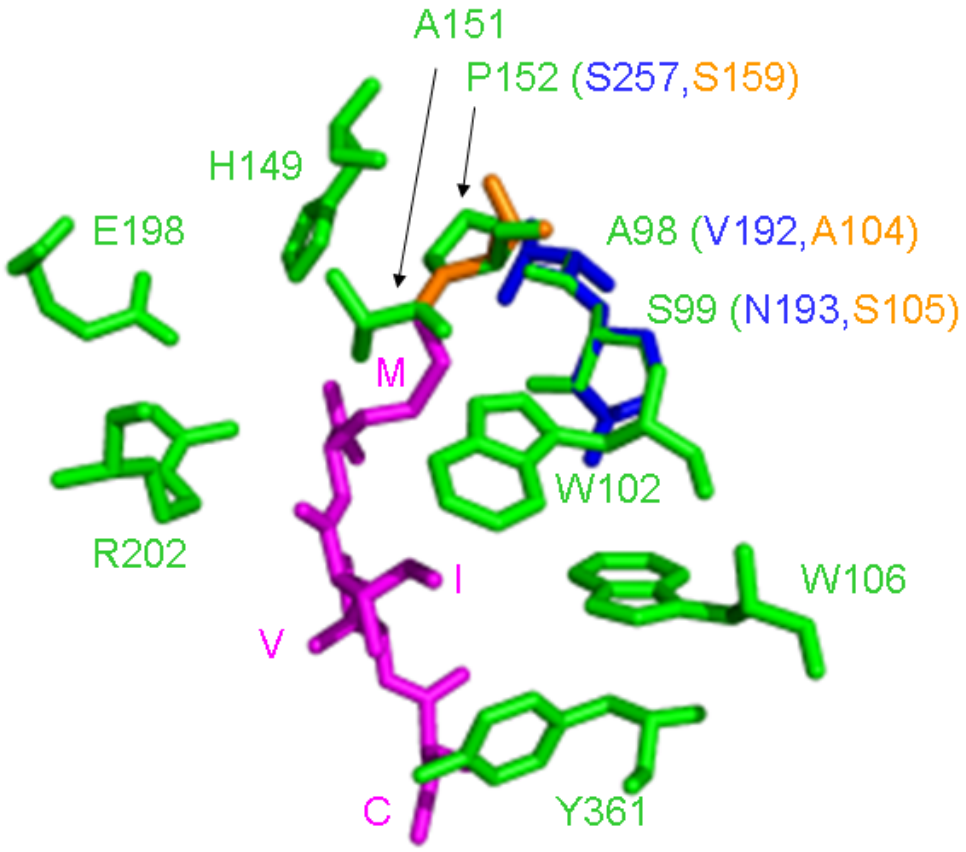


Figure 4.13 Structural comparison between β-subunits of rPFTase, CaPFTase and yPFTase in (a) linear sequence and in (b) 3-D structure. (a) Sequence alignment of three different PFTases. The residues that are close to the Ca_1a_2X substrate are indicated by arrows. The multiple sequence alignment was done using ClustalW from the ExPASy website. (b) CVIM peptide binding in PFTase. Only the substrate peptide Ca_1a_2X motif and protein residues involved in binding are shown. The structure was derived from the rPFTase crystal structure (PDB ID: 1D8D). If the corresponding residues are different in CaPFTase or yPFTase, those residues are superimposed in different colors. Residues from rPFTase, CaPFTase and yPFTase are shown in green, blue and orange, respectively. The CVIM peptide substrate is shown in magenta.

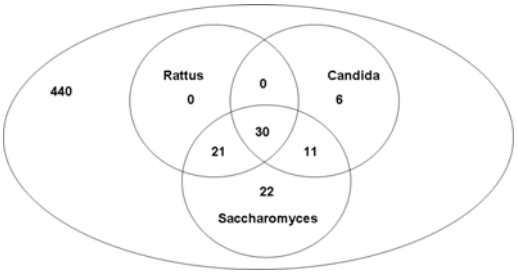
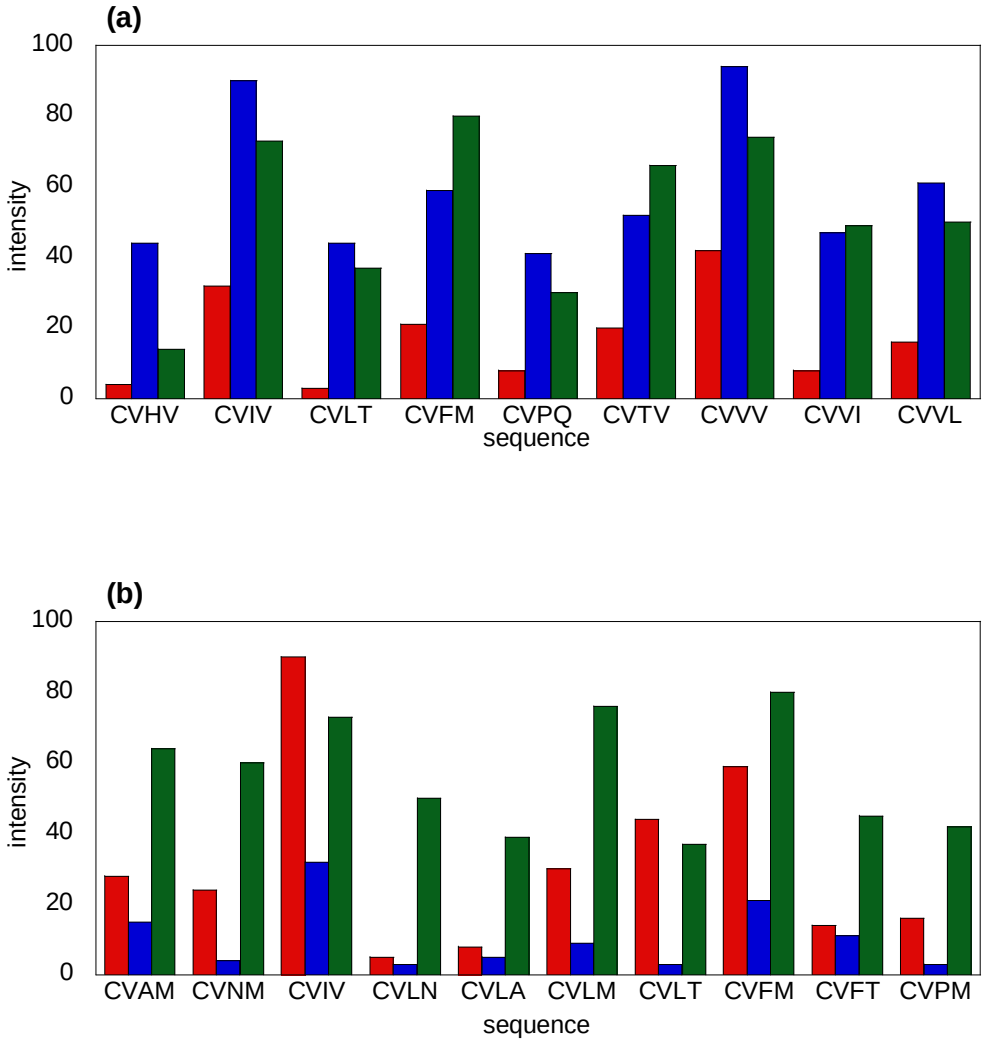


Figure 4.14 Number of peptide sequences that are recognized by three different PFTases. Peptides which showed high and medium intensities in the screening are grouped together. Low intensity spots were not considered in this analysis.



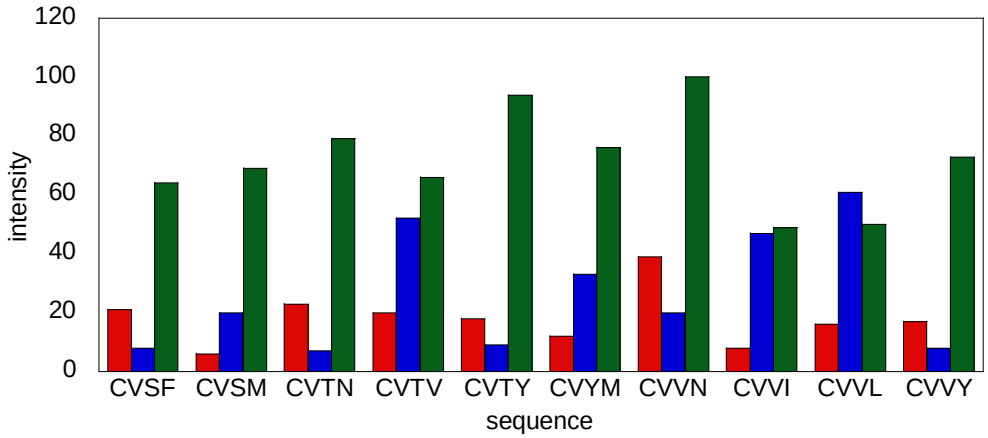


Figure 4.15 (a) Comparison of the intensities of selected peptide sequences which show at least a 2-fold higher intensity with CaPFTase relative to rPFTase from Figure 4.10. For each sequence, the left bar indicates the iscreening intensities by using the PFTase from *R. norvegicus*, the middle bar from *C. albicans* and the right bar from *S. cerevisiae*. (b) Comparison of the intensities of selected peptide sequences which show at least a 2-fold higher intensity with yPFTase relative to rPFTase from Figure 4.11. For each sequence, the left bar indicates the iscreening intensities by using the PFTase from *R. norvegicus*, the middle bar from *C. albicans* and the right bar from *S. cerevisiae*.

Ca_1a_2X	Enzyme	Screening intensity [a]	k_{cat} (s^{-1})	K_M (uM)[b]	k_{cat} / K_M ($uM^{-1}s^{-1}$)[b]
CCVQ	rPFTase	82 (or 106)	$(7.3\pm0.7)\times10^{-2}$	0.89±0.30	0.082
CCIA		52 (or 33)	NA[c]	NA[c]	NA[c]
CVIA	rPFTase	73	0.14±0.01	0.74±0.29	0.19
CVIT	rPFTase	66	0.14±0.01	1.2±0.3	0.12
CVIN	rPFTase	53	0.16±0.01	1.5±0.3	0.11
CTIL	rPFTase	21	0.13±0.01	5.6±1.0	0.023
	CaPFTase	84	1.3(±0.2)	6.1±1.6	0.21
CVVV	rPFTase	42	$(7.8\pm0.9)\times10^{-2}$	0.63±0.37	0.12
	CaPFTase	94	1.6±0.1	2.1±0.4	0.74
	yPFTase	74	2.4±0.3	4.4±1.4	0.55
CVVN	rPFTase	39	0.12±0.01	3.4±0.5	0.035
	CaPFTase	20	1.4±0.1	8.6±1.4	0.16
	yPFTase	100	1.5±0.1	1.7±0.4	0.88
CVTN	rPFTase	23	0.25±0.08	15±6	0.016
	CaPFTase	7	1.9±0.3	49±9	0.039
	yPFTase	79	2.4±0.1	14±1	0.17

[a] The intensity was normalized to CIIS. The number in parentheses was obtained when a lower concentration of rPFTase was used.

[b] K_M in this case is the K_M for the peptide substrate.

[c] The reaction rate did not change much during the range of concentration of DsGCCIA and hence cannot be fit into Michaelis-Menten equation. But at 2 uM peptide concentration and 4 uM C15-Alk, the rate of DsGCCIA is 5.3 times slower than DsGCCVQ.

Table 4.3 Evaluation of Ds-GCaaX peptides for *R. norvegicus* or *C. albicans* PFTase. Assays were performed at varying concentration of dansylated peptides (1-8 μM), 50-100 nM PFTase, 4 μM C15-Alk, 50 mM Tris, pH 7.5, 5 mM DTT, 5 mM $MgCl_2$, 50 μM $ZnCl_2$, and 0.040 % (w/v) n-dodecyl-ß-D-maltoside.

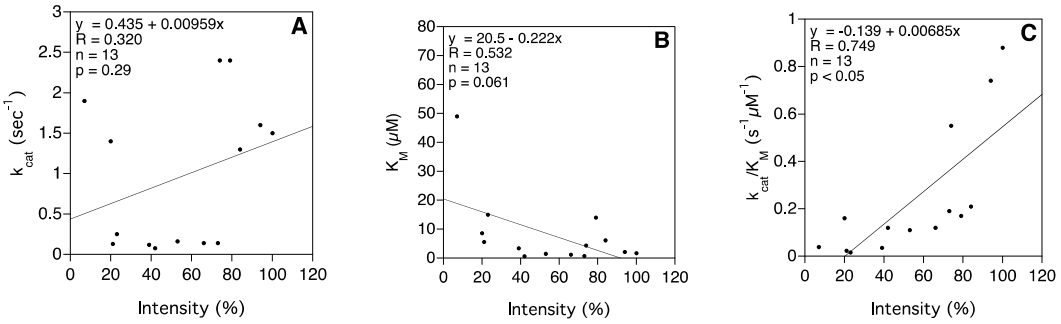


Figure 4.16 Plot of spot intensity versus (a) k_{cat}, (b) K_M, (c) k_{cat}/K_M from Table 4.3. The first two peptides data from Table S4 were excluded because their spot intensities were normalized differently than the others. The linear fit equation, R value and p values are shown as insets.

4.8 Conclusion

In summary, two 380-member peptide libraries (CVa_2X and CCa_2X libraries) were evaluated as substrates for *R. norvegicus* PFTase. That screening revealed a number of both known and unknown substrate sequences. Some of the latter occur in the genomes of bacteria and viruses and may be important for pathogenesis, suggesting new potential targets for therapeutic intervention. Screening of the CVa_2X library with alkyne-functionalized isoprenoid substrates showed that those prepared from C_{10} or C_{15} precursors gave similar results while the analogue synthesized from a C_5 unit gave a different pattern of reactivity. Finally the substrate specificities of PFTases from three organisms (*S. cerevisiae*, *R. norvegicus* and *C. albicans*) were compared using CVa_2X libraries. *R. norvegicus* PFTase was found to share more peptide substrates with *S. cerevisiae* PFTase than with *C. albicans* PFTase. Overall, the 380-member CVa_2X library was screened using three different isoprenoid

substrates and three different enzymes while the CCa_2X library was screened using two different enzyme concentrations; thus over 3000 different combinations were evaluated. In general, this is a highly efficient strategy for rapidly probing the specificity of this important enzyme and should be useful for a variety of future studies on PFTase and related enzymes in the future.

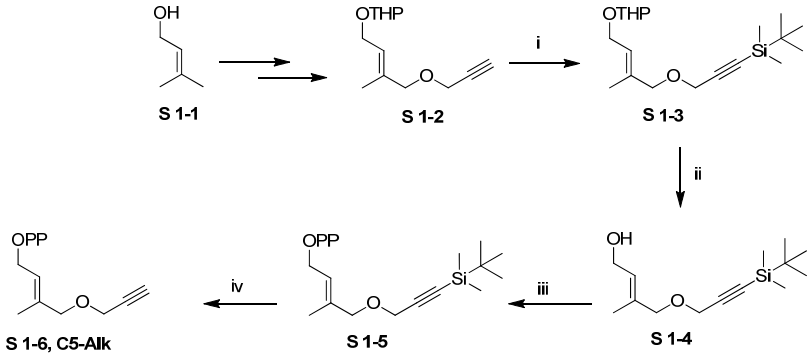
4.9 Materials and methods

(a) General

All synthetic reactions were carried out at 25 ^{o}C unless otherwise noted. Deuterated NMR solvents were purchased from Cambridge Isotope Laboratories, Inc. Biotin-azide was purchased from ChemPep Inc. All amino acids and peptide coupling reagents were purchased from P3 BioSystems or NovaBiochem. Other general reagents, including salts and reagents, were purchased from Sigma-Aldrich. ^{1}H NMR spectra were obtained at 300 MHz; ^{31}P NMR spectra were obtained at 121 MHz. All NMR spectra were acquired on Varian instruments at 25 °C. Chemical shifts are reported in ppm and J values are in Hz. SPOT synthesis was performed using a ResPep SL machine obtained from INTAVIS Bioanalytical Instruments. Fluorescence assay data were obtained using a Varian Cary Eclipse Fluorescence Spectrophotometer. Analytical HPLC was performed with a Beckman model 125/166

instrument, equipped with a diode array UV detector, ABI Analytical Spectroflow 980 fluorescence detector, and a Varian C18 column (Microsorb-MV, 5 μm, 4.6 x 250 mm). Preparative HPLC separations were performed by using a Beckman model 127/166 instrument, equipped with a UV detector and a Phenomenex C18 column (Luna, 10 μm, 10 x 250 mm). MS spectra for small molecules were obtained on a Bruker BioTOF II instrument. MS spectra for peptide libraries were obtained on Bruker Biflex III instrument.

(b) Synthesis of C5-Alk:



Scheme S4.1. Synthesis of C5-Alk. Reagents and conditions: (i) n-BuLi, -78 ^{o}C, 1 h, then TBDMS, RT, 4 h, THF; (ii) PPTS, CH_3OH, RT, 18 h; (iii) PBr_3, Et_2O, 0 ^{o}C, 45 min, then $(nBu_4N)_3HP_2O_7$, CH_3CN, RT, 16 h; (iv)TBAF, CH_3CN, RT, 4 h.

Compounds **S1-2** was made by following a previously published procedure.[136]

Compound **S1-3**

To a solution of **S1-2** (5.06 mmol) in THF (5 mL) which was pre-cooled to -78 ^{o}C was added n-BuLi (2.5 mL, 2.5 M, 6.25 mmol) dropwise. The mixture was stirred

at -78 ^{o}C for 1 h. A solution of TBDMSCl (0.994 g, 6.58 mmol) in THF (2.5 mL) was added to the above solution dropwise at -78 ^{o}C. The mixture was stirred at RT for 4 h. The reaction was quenched by adding H_2O (25 mL). The solution was extracted with Et_2O (2 X 25 mL) and the combined organic layers were dried with $MgSO_4$ and concentrated by rotary evaporation to yield **S1-3** as a colorless oil. Crude compound (**S1-3**) was used in the next step without further purification.

Compound **S1-4**

To a solution of **S1-3** (~5.06 mmol) in CH_3OH (10 mL) was added PPTS (0.175 g, 0.759 mmol). The mixture was stirred at RT for 18 h. The solvent was removed by rotary evaporation and the resulting oil was purified by silica gel chromatography (Et_2O/hexanes = 2/3) to yield **S1-4** as a colorless oil (0.983 g, 76% in two steps). ^{1}H NMR ($CDCl_3$): δ 5.69 (t, J = 7.5, 1H), 4.22 (dd, J = 6, 6, 2H), 4.14 (s, 2H), 3.97 (s, 2H), 1.71 (s, 3H), 1.59 (s, 3H), 0.94, (s, 9H), 0.11 (s, 6H). ESI-MS [M+H^{+}]: m/z calc. for $C_{14}H_{26}NaO_2Si^{+}$: 277.1600; found: 277.1608.

Compound **S1-5**

To a solution of **S1-4** (73 mg, 0.287 mmol) in Et_2O (4 mL) which was pre-cooled to 0 ^{o}C was added PBr_3 (30 uL, 0.316 mmol). The mixture was stirred at 0 ^{o}C for 55

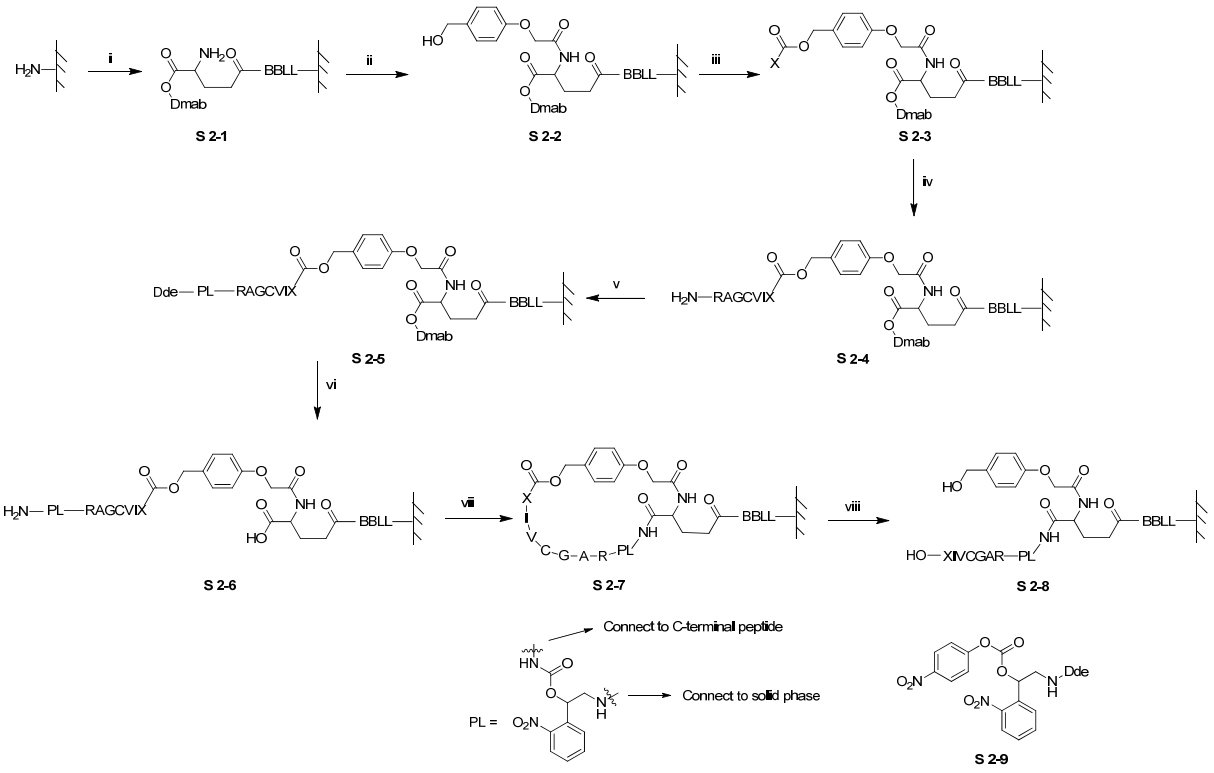
min. The reaction was quenched by adding a few drops of H_2O at 0 ^oC. Cold saturated $NaHCO_3$ (15 mL) was then added. The mixture was extracted with cold Et_2O (2 X 15 mL) and the combined organic layers were dried with $MgSO_4$ and concentrated by rotary evaporation to yield a colorless oil which was then dissolved in CH_3CN (4 mL). Next, $(nBu_4N)_3HP_2O_7$ (0.514 g, 0.574 mmol) was added to this solution. The mixture was stirred at RT for 16 h. A column was filled with Dowex 50WX8 ion-exchange resin which was converted to the ammonium form by adding three column volumes of H_2O/ 30% NH_4OH (3:1, v/v) followed by equilibration with buffer C (25 mM NH_4HCO_3/iPrOH 49:1, v/v). The reaction was loaded onto the column and eluted with buffer C. The combined fractions were lyophilized and purified by preparative RP-HPLC with linear gradient from 25 mM NH_4HCO_3 to 35% CH_3CN over 50 min. The fractions containing **S1-5** were collected and lyophilized. ^1H NMR (D_2O/ND_4OD): δ 5.54 (t, J = 6, 1H), 4.35 (dd, J = 7.5, 7.5, 2H), 4.02 (s, 2H), 3.88 (s, 2H), 1.54 (s, 3H), 0.76 (s, 9H), -0.04 (s, 6H). ^{31}P NMR (D_2O/ND_4OD): δ -5.83 (d, J = 22.1, 1P), -9.83 (d, J = 22.1, 1P). ESI-MS $[M-H]^-$: m/z calc. for $C_{14}H_{27}O_8P_2Si^-$: 413.0956; found: 413.0943.

Compound **S1-6**

Solid **S1-5** from above was suspended in CH_3CN (4 mL) and TBAF (200 uL, 1M,

0.2 mmol) was added to the solution. The mixture was stirred at RT for 190 min. 25 mM NH_4HCO_3 (1 mL) was added to quench the reaction. A column was filled with Dowex 50WX8 ion-exchange resin which was converted to the ammonium form by adding three column volumes of H_2O/ 30% NH_4OH (3:1, v/v) followed by equilibration with buffer C (25 mM NH_4HCO_3/iPrOH 49:1, v/v). The reaction was loaded onto the column. The column was eluted with buffer C. The combined fractions were lyophilized and purified by preparative RP-HPLC with 100% 25 mM NH_4HCO_3. The fractions containing **S1-6** were collected and lyophilized. ^{1}H NMR (D_2O/ND_4OD): δ 5.56 (t, J = 7.5, 1H), 4.35 (dd, J = 7.5, 7.5, 2H), 4.01 (s, 2H), 3.88 (s, 2H), 1.54 (s, 3H). ^{31}P NMR (D_2O/ND_4OD): δ -5.82 (d, J = 22.7, 1P), -9.82 (d, J = 22.1, 1P). ESI-MS $[M-H]^-$: m/z calc. for $C_8H_{14}O_8P_2^-$: 299.0091; found: .299.0094.

(c) Synthesis of inverted peptides:



Scheme S4.2. Synthesis of inverted peptides for screening assay. Reagents and conditions: (i) standard DIC coupling (2X), then, capping, then 20% piperidine; (ii) standard DIC coupling (2X); (iii) 0.4 M Fmoc-Aa and 1.2 M CDI in DMF (4X), then, capping, then 20% piperidine; (iv) standard DIC coupling (2X), then, capping, then 20% piperidine; (v) 0.5 M photocleavable linker, 0.5 M Et_3N in DMF (3X); (vi) 2% NH_2NH_2; (vii) 0.05 M BOP, 0.05 M 6-Cl-HOBt and 0.1 M DIEA in DMF (2X); (viii) modified reagent K.

SPOT Synthesis

Amino-functionalized membranes were either purchased from INTAVIS or made from Whatman 540 paper.[2] The TOTD (trioxa-tridecanediamine) membranes[2] were coupled to β-Ala before usage. A 10×15 cm^2 TOTD membrane was incubated with 0.893 g Fmoc-β-Ala (2.87 mmol), 0.488 g 6-Cl-HOBt (2.87 mmol), and 443 uL DIC (2.87 mmol) in 20 mL DMF for 105 min. The membrane was washed with DMF 1X.

The membrane was incubated with 2% (v/v) AcOH in DMF for 5 min, then incubated with 2% (v/v) AcOH and 2% (v/v) DIEA in DMF for 15 min. The membrane was washed with DMF 4X. The β-Ala loading was 0.19(±0.02) umol/cm^2 based on Fmoc assay. The membrane was incubated with 20% piperidine in DMF for 5 min 2X. The membrane was washed with DMF 4X, CH_3OH 2X, and Et_2O 2X.

The peptides were synthesized using a ResPep SL SPOT-robot (INTAVIS Bioanalytical Instruments). The synthesis started with spot definition by a standard protocol followed by the coupling of a solution of 0.11 μL of Fmoc-amino acid (0.5 M), 0.055 μL of DIC (1.1 M) and 0.055 μL of 6-Cl-HOBt (1.2 M) in DMF (double coupling, 10 and 15 min reaction each). The membrane was capped with Ac_2O in DMF (2%), washed with DMF (3X) and EtOH (3X) and finally air-dried. Fmoc was removed by 20% piperidine in DMF (3X). After HMPA was coupled to the growing peptide chain using the above procedure, the membrane was washed with DMF (3X), EtOH (3X) without capping.

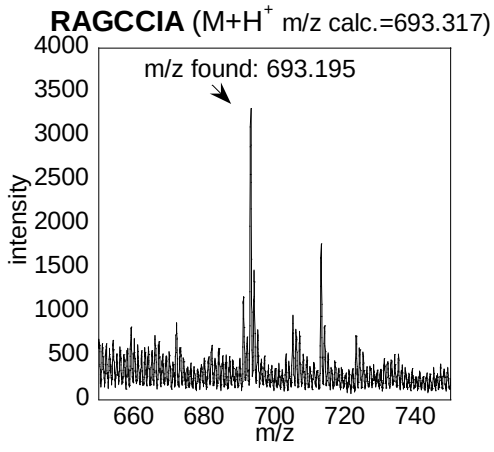
For ester-bond formation (step iii), 0.22 μL of solution Fmoc-amino acid (0.4M) activated with 1,1'-carbonyldiimidazole (CDI, 3 equiv.) in DMF were spotted on the membrane (4X coupling, 15 min reaction time each). CDI was substituted by CDT if the amino acid was Gln or Tyr.

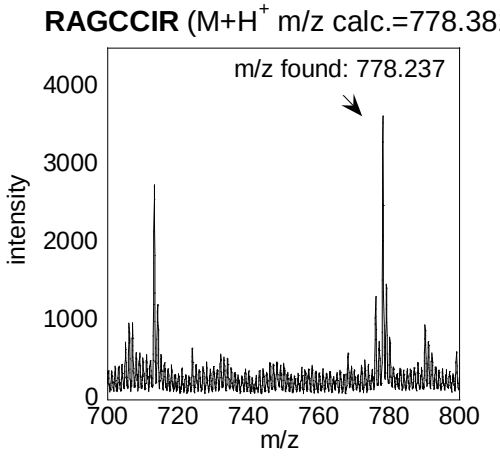
To add the photocleavable linker, 0.22 μL of a DMF solution containing **S 2-9**

(0.5 M) and Et_3N (1.0 M) were spotted onto the membrane (3X coupling, 15 min reaction time each).

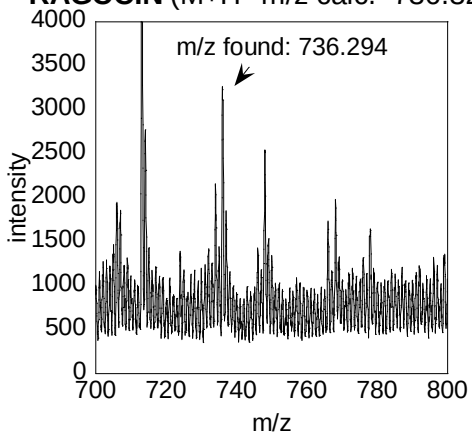
For peptide cyclization, the membrane was incubated with a solution of (BOP)/6-Cl-HOBt/DIEA (0.05 M, 0.05 M, 0.1 M respectively) in DMF. This reaction was performed twice, the first time for 2 h and the second time for 16 h. The membrane was washed with DMF (3X), CH_3OH (2X) and Et_2O (2X).

The membrane was treated with a modified reagent K (6.5% phenol, 5% H_2O, 5% thioanisole, 2.5% ethanedithiol, 1% anisole, and 1% triisopropylsilane in TFA) for 2 h for ester cleavage (ring opening) and global side-chain deprotection.

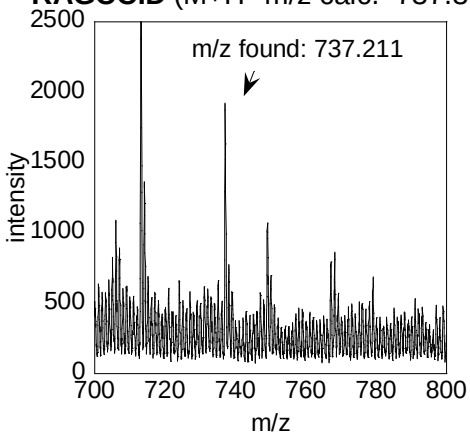




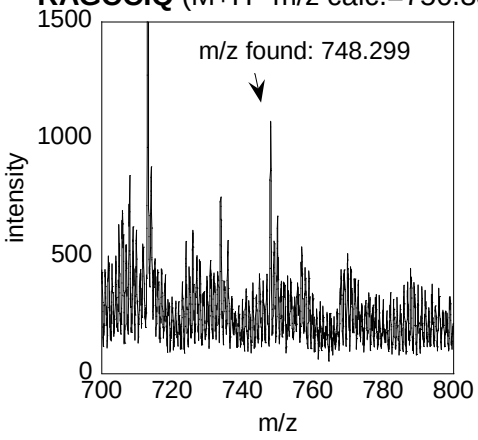
RAGCCIN ($M+H^+$ m/z calc.=736.323)



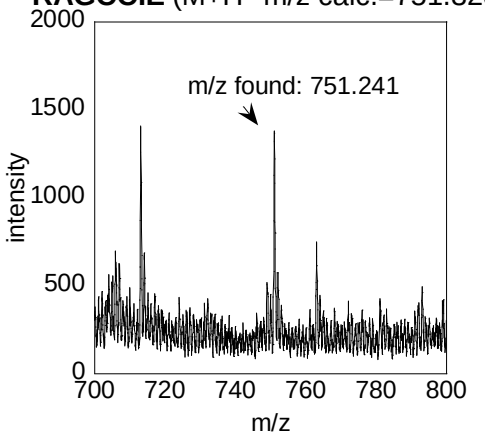
RAGCCID ($M+H^+$ m/z calc.=737.307)



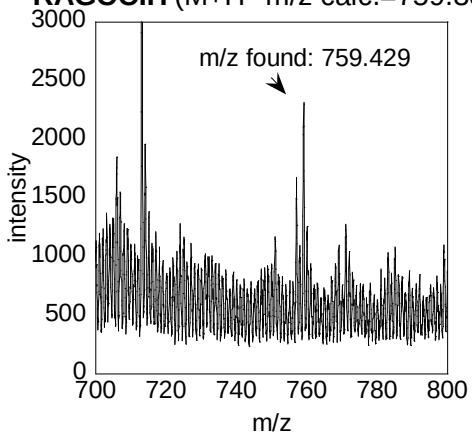
RAGCCIQ ($M+H^+$ m/z calc.=750.339)



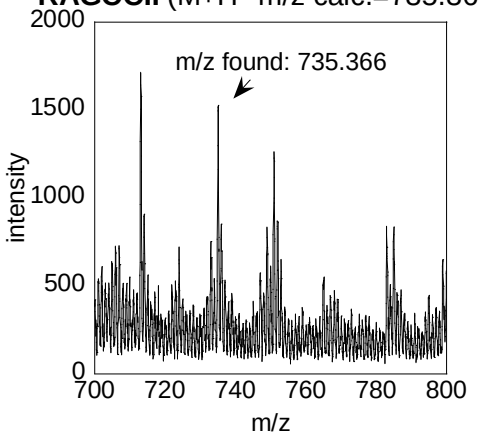
RAGCCIE ($M+H^+$ m/z calc.=751.323)



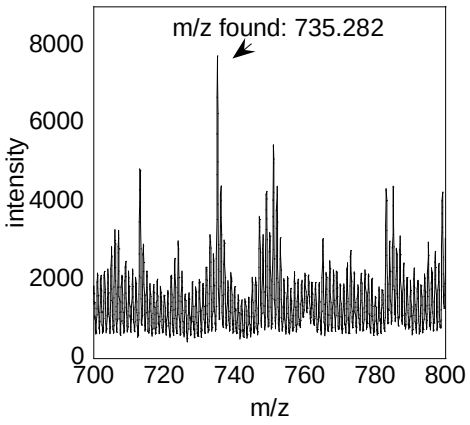
RAGCCIH ($M+H^+$ m/z calc.=759.339)



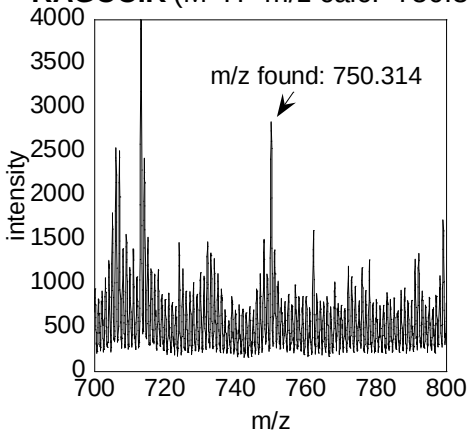
RAGCCII ($M+H^+$ m/z calc.=735.364)



RAGCCIL ($M+H^+$ m/z calc.=735.364)

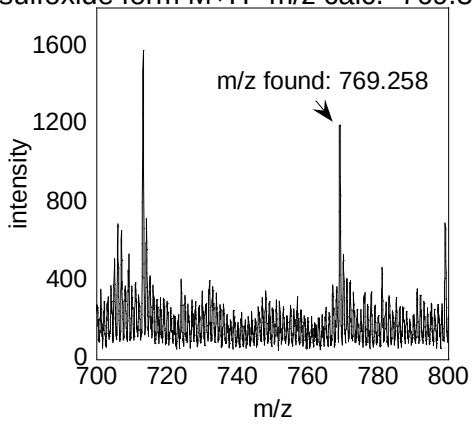


RAGCCIK ($M+H^+$ m/z calc.=750.375)

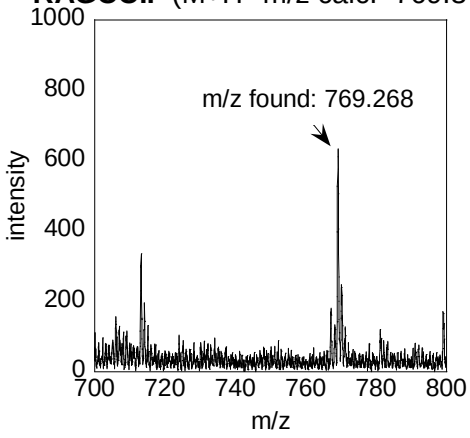


RAGCCIM ($M+H^+$ m/z calc.=753.320)

(sulfoxide form $M+H^+$ m/z calc.=769.315)

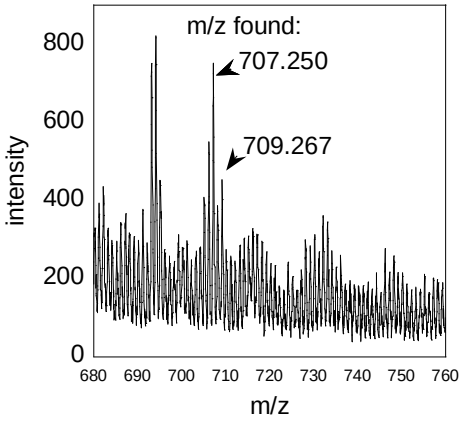


RAGCCIF ($M+H^+$ m/z calc.=769.348)

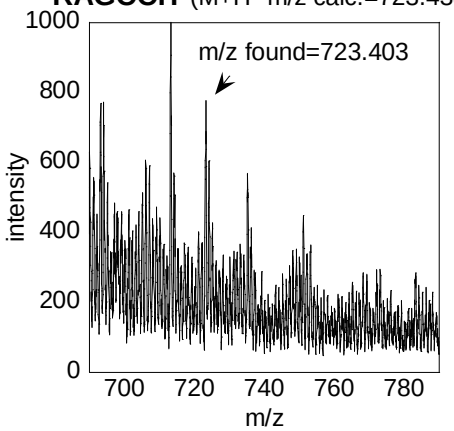


RAGCCIS ($M+H^+$ m/z calc.=709.312)

disulfide form $M+H^+$ m/z calc.=707.296



RAGCCIT ($M+H^+$ m/z calc.=723.438)



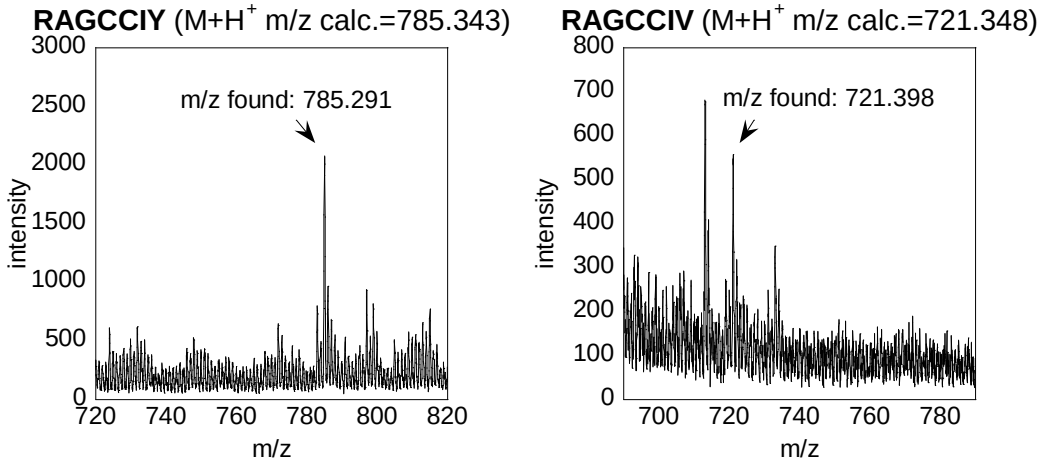


Figure S4.1 MALDI analysis of peptides produced by SPOT synthesis performed to verify the production of the desired product. The photocleavable moiety shown in Scheme S2 was used as the linker. The individual spot was treated under UV lamp (365 nm) in CH_3CN/H_2O (v/v = 5/95) for 2 h. The spots were washed with CH_3CN and CH_3OH. The combined solvents were removed *in vacuo*. The peptides (RAGCCIX, X≠P) were re-dissolved in CH_3CN/H_2O (v/v = 50/50) with 0.1% TFA. The m/z shown was manually calibrated by an internal standard (3XCCA+H^+). Peptides ending with CCIC, CCIG, CCIW either had low signal intensity or cannot be found.

(d) Screening for PFTase specificity:

For experiments using C15-Alk, the whole membrane (10X15 cm^2) was incubated with 155 mM DTT for 1 h and then washed with H_2O (1X). Typical conditions for PFTase reactions were 5.1 mM DTT, 5 mM $MgCl_2$, 50 μM $ZnCl_2$, 50 mM Tris-HCl (pH=7.5), 1 mg/mL BSA, ~100 μM C5-Alk, 2.7 ug/mL *S. cerevisiae* PFTase (or 55 ug/mL *R. norvegicus* PFTase, or 4.2 ug/mL *C. albicans* PFTase). The enzymatic reactions were typically allowed to proceed for 5 h followed by washing with 25 mM NH_4HCO_3 (3X), CH_3CN (3X), and H_2O (3X).

The CCa_2X library membrane was enzymatically prenylated with 55 ug/mL or

182 ug/mL rPFTase for 5.5 h using the same procedure as described above.

For experiments using C10-Alk, typical conditions for PFTase reactions were 5.1 mM DTT, 5 mM $MgCl_2$, 50 μM $ZnCl_2$, 50 mM Tris-HCl (pH=7.5), 1 mg/mL BSA, ~100 μM FPP analogue, 40 ug/mL rPFTase. The enzymatic reactions were typically allowed to proceed for 5 h followed by washing with 25 mM NH_4HCO_3 (3X), CH_3CN (3X), and H_2O (3X).

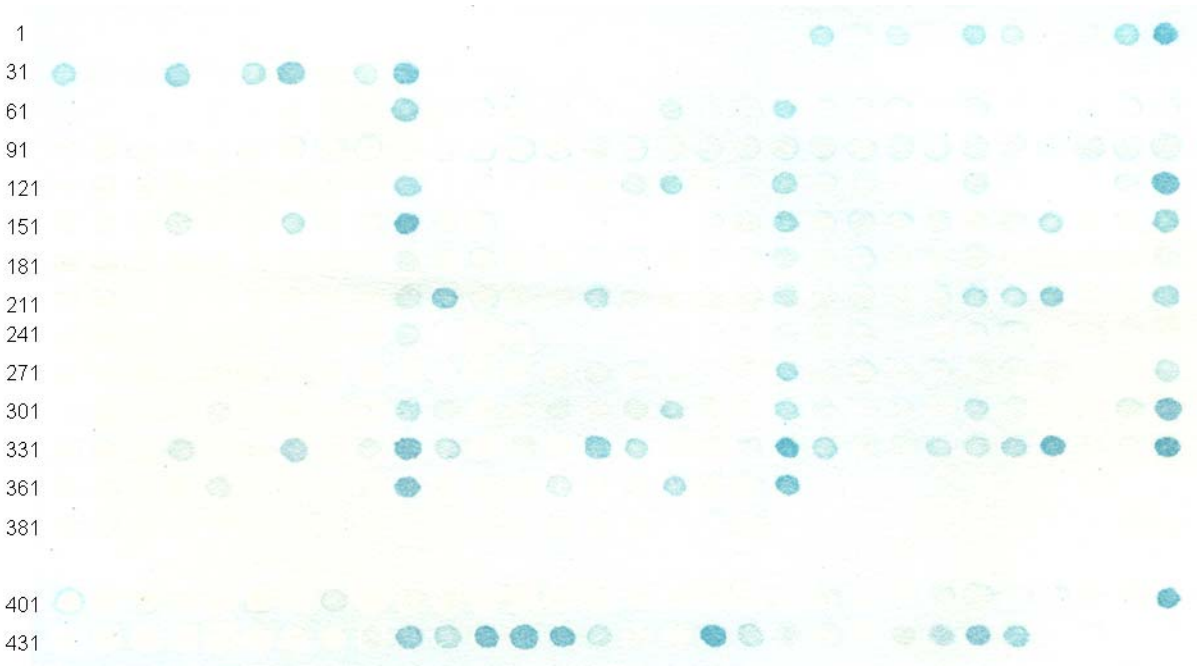
For experiments using C5-Alk, typical conditions for PFTase reactions were 5.1 mM DTT, 5 mM $MgCl_2$, 50 μM $ZnCl_2$, 50 mM Tris (pH=7.5), 1 mg/mL BSA, ~100 μM FPP analogue, 55 ug/mL rPFTase. The enzymatic reactions were allowed to proceed for 7.5 h followed by washing with 25 mM NH_4HCO_3 (3X), CH_3CN (3X), and H_2O (3X).

Typical conditions for click reactions were 0.1 mM biotin-azide, 1 mM TCEP, 0.2 mM TBTA, 1 mM $CuSO_4$ in PBS buffer 1 (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 2 mM KH_2PO_4, pH=7.4). The reaction was typically allowed to proceed for 16 h. After reactions, the membranes were washed with H_2O (3X), DMF (3X), and H_2O (3X).

For detection of prenylated proteins, the membranes were incubated with 5 % milk in PBS buffer 2 (100 mM KH_2PO_4, 500 mM NaCl, pH=6.5) for ~35 min and then incubated with SP-AP in 5% milk PBS buffer 2 (v/v = 1/800) for ~40 min. After

reactions, the membranes were washed with H_2O (9X). Upon addition of BCIP solution (0.313 mg/mL, 20 mL) in buffer E (5 mM $MgCl_2$, 20 μM $ZnCl_2$, 100 mM NaCl, 30 mM Tris, pH=8.5), turquoise color developed on positive spots after several hours depending on the specific library or isoprenoid. The staining reaction was terminated by washing with H_2O (2X) and CH_3CN (1X).

The quantification of color intensity was performed using Image J software. The membrane was scanned and saved. The file was opened using Image J. Under Image>Type, 8-bit was selected to convert the image to grayscale. Under the menu Process>Subtract Background, the rolling ball radius was set to 50. Under Analyze>Set Measurements, the Area, Mean Gray Value, and Integrated Density were selected. Under Analyze>Set Scale, “pixels” in the box next to Unit of length was selected. Under Edit>Invert, the colors on the image were inverted, a line was drawn around the image boundary and the measurement of the selected spots was then performed using the “m” key.


1
31
61
91
121
151
181
211
241
271
301
331
361
381
401
431

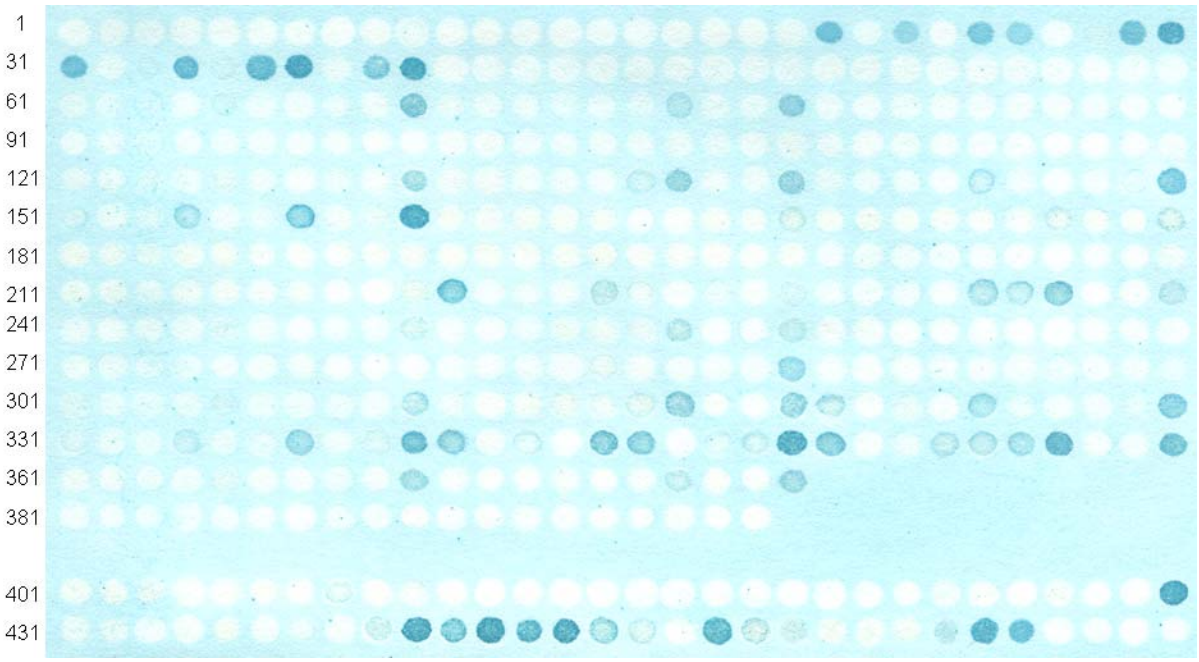

1
31
61
91
121
151
181
211
241
271
301
331
361
381
401
431

Figure S4.2 Screening of enzymatic prenylation of an RAGCVa_2X library. Each membrane was prenylated with OPP-C15-alkyne by rPFTase (55 ug/mL), clicked with biotin-azide and then visualized by SP-AP in BCIP solution.

Spot identity:

1-20: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)E

21-40: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)Q

41-60: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)D

61-80: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)N

81-100: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)R

101-120: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)K

121-140: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)H

141-160: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)A

161-180: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)V

181-200: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)I

201-220: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)L

221-240: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)F

241-260: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)Y

261-280: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)W

281-300: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)G

301-320: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)C

321-340: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)M

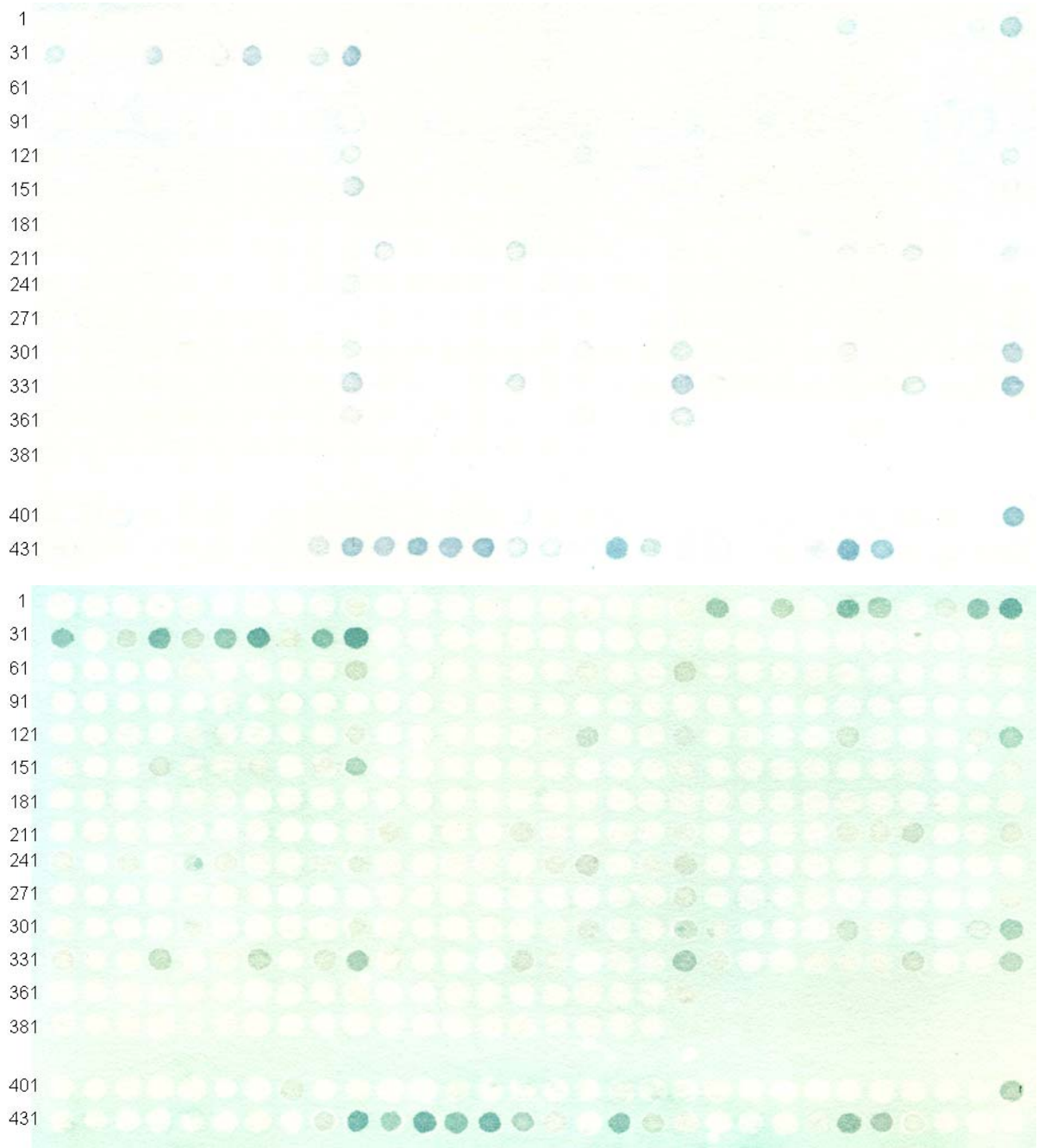
341-360: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)S

361-380: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)T

381-399: SVI(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,S,T,W,Y,V)

401-410: CTRK, CQRK, CKCI, CSEI, CIII, CKYI, CASL, CTIL, CSGL, CAIL
411-420: CIIL, CIDL, CPFW, CHDE, CRME, CKGE, CCLD, CHHD, CDDD, CALD
421-430: CWKD, CDPN, CRNR, CQGK, CLAK, CRVK, CIGK, CIHH, CYNA, CTVA
431-440: CSNA, CPKA, CYDA, CNDV, CMYV, CFIF, CIQF, CNAG, CCVC, CSIM
441-450: CCIM, CTIM, CIIM, CAIM, CTLM, CCPS, CDFS, CIIS, CCCS, CKQS
451-460: CIKS, CTDS, CKCT, CKQQ, CCIQ, CASQ, CDDY, CAPY, CADY, CPNY

(a)



(b)

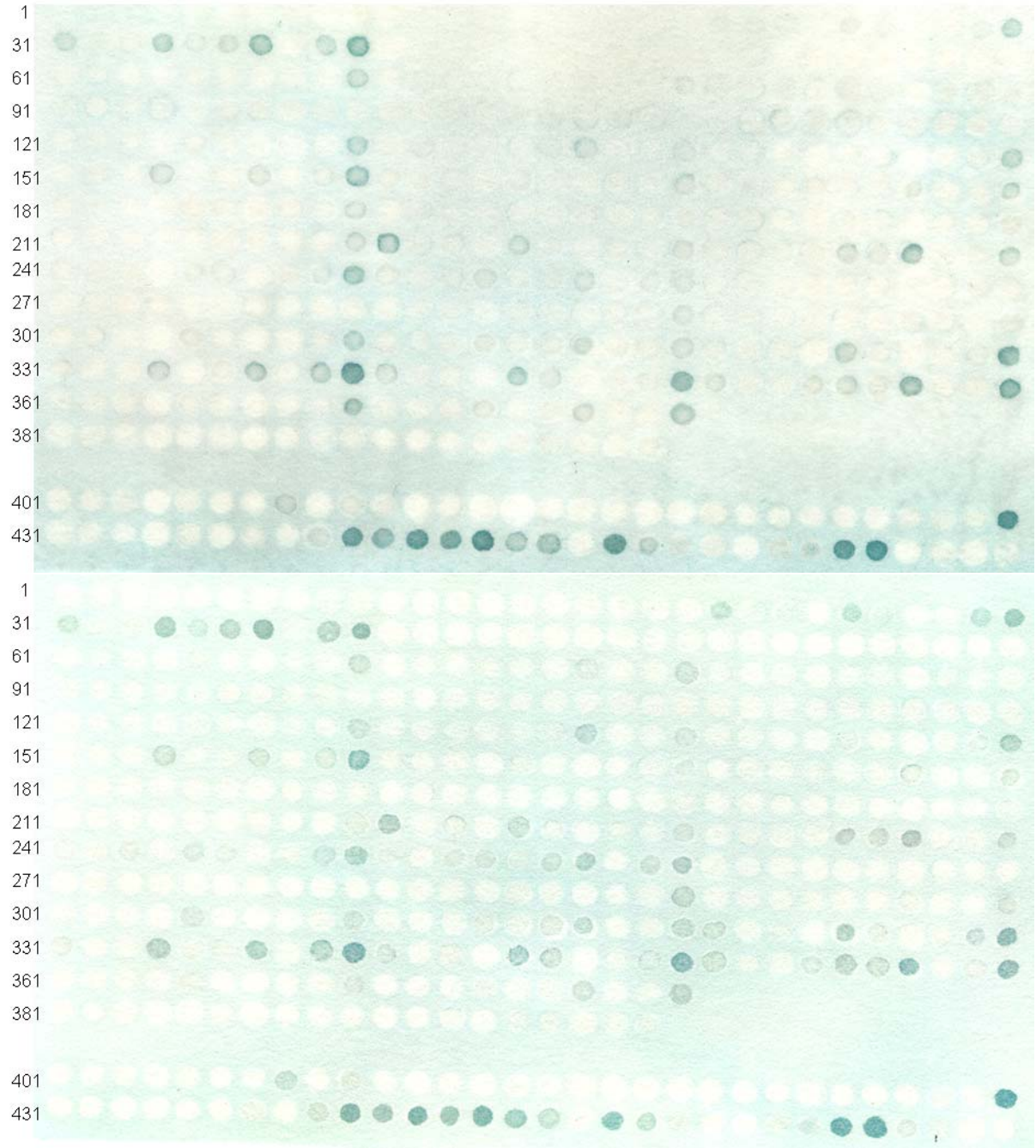


Figure S4.3 Screening of enzymatic prenylation of an RAGCCa_2X library. Each membrane was prenylated with OPP-C15-alkyne by rPFTase (55 ug/mL in (a), 182 ug/mL in (b)), clicked with biotin-azide and then visualized by SP-AP in BCIP solution.

Spot identity:

1-20: CC(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)E

21-40: CC(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)Q

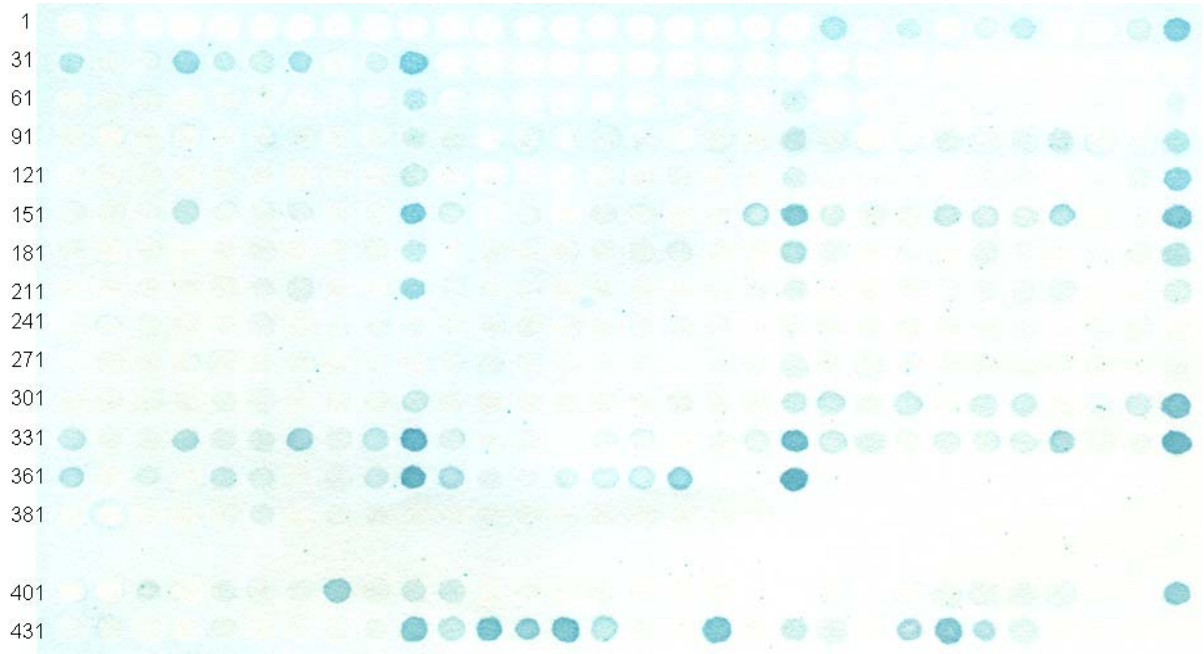
41-60: CC(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)D

61-80: CC(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)N

81-100: CC(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)R

101-120: CC(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)K

121-140: CC(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)H
141-160: CC(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)A
161-180: CC(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)V
181-200: CC(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)I
201-220: CC(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)L
221-240: CC(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)F
241-260: CC(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)Y
261-280: CC(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)W
281-300: CC(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)G
301-320: CC(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)C
321-340: CC(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)M
341-360: CC(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)S
361-380: CC(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)T
381-399: SCI(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,S,T,W,Y,V)
401-410: CTRK, CQRK, CKCI, CSEI, CIII, CKYI, CASL, CTIL, CSGL, CAIL
411-420: CIIL, CIDL, CPFW, CHDE, CRME, CKGE, CCLD, CHHD, CDDD, CALD
421-430: CWKD, CDPN, CRNR, CQGK, CLAK, CRVK, CIGK, CIHH, CYNA, CTVA
431-440: CSNA, CPKA, CYDA, CNDV, CMYV, CFIF, CIQF, CNAG, CCVC, CSIM
441-450: CCIM, CTIM, CIIM, CAIM, CTLM, CCPS, CDFS, CIIS, CCCS, CKQS
451-460: CIKS, CTDS, CKCT, CKQQ, CCIQ, CASQ, CDDY, CAPY, CADY, CPNY



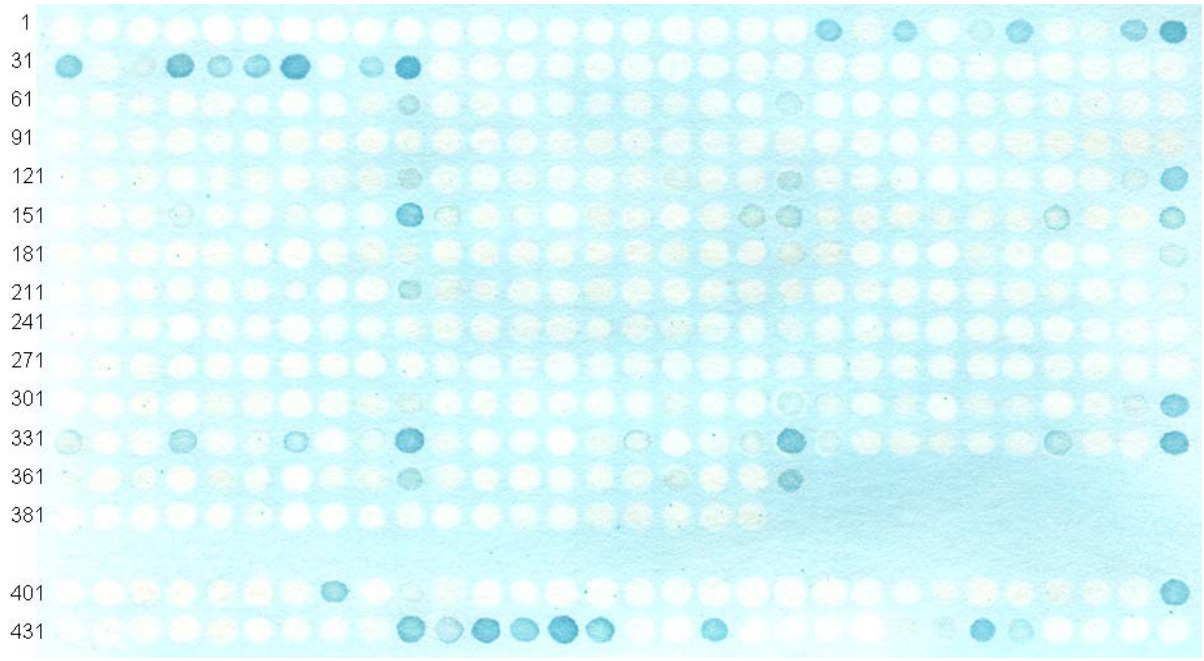


Figure S4.4 Screening of enzymatic prenylation of an RAGCVa_2X library. Each membrane was prenylated with OPP-C15-alkyne by CaPFTase (4.2 ug/mL), clicked with biotin-azide and then visualized by SP-AP in BCIP solution.

Spot identity:

1-20: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)E

21-40: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)Q

41-60: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)D

61-80: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)N

81-100: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)R

101-120: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)K

121-140: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)H

141-160: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)A

161-180: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)V

181-200: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)I

201-220: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)L

221-240: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)F

241-260: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)Y

261-280: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)W

281-300: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)G

301-320: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)C

321-340: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)M

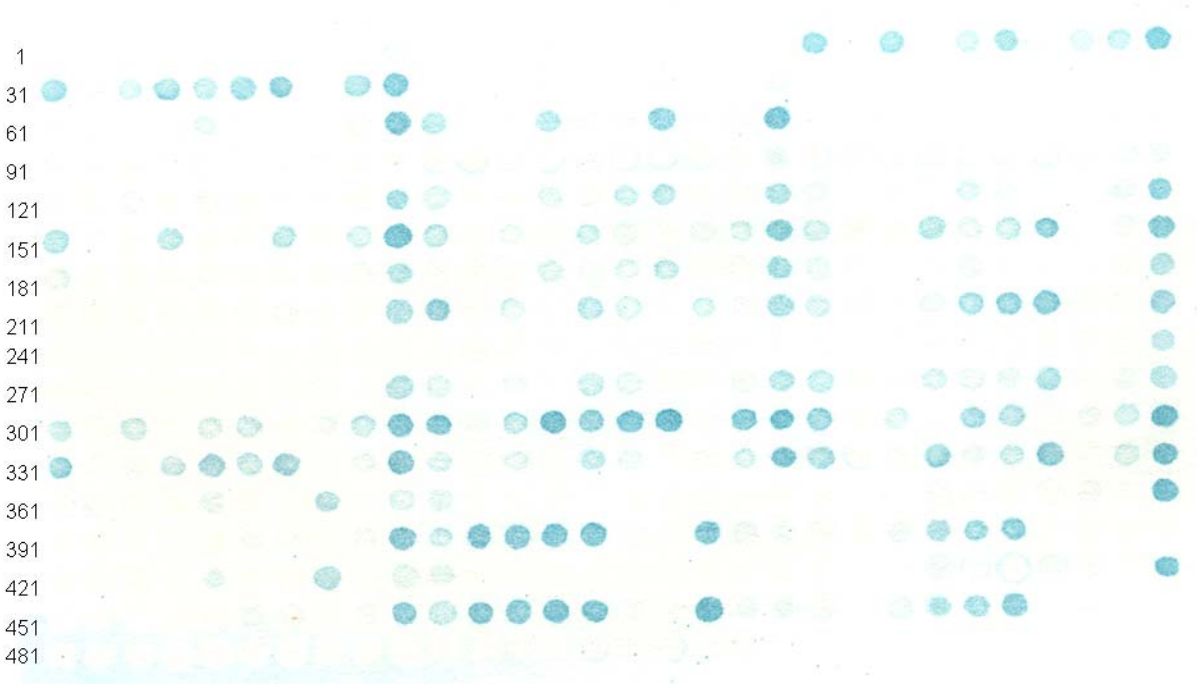
341-360: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)S

361-380: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)T

381-399: SVI(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,S,T,W,Y,V)

401-410: CTRK, CQRK, CKCI, CSEI, CIII, CKYI, CASL, CTIL, CSGL, CAIL
411-420: CIIL, CIDL, CPFW, CHDE, CRME, CKGE, CCLD, CHHD, CDDD, CALD
421-430: CWKD, CDPN, CRNR, CQGK, CLAK, CRVK, CIGK, CIHH, CYNA, CTVA
431-440: CSNA, CPKA, CYDA, CNDV, CMYV, CFIF, CIQF, CNAG, CCVC, CSIM
441-450: CCIM, CTIM, CIIM, CAIM, CTLM, CCPS, CDFS, CIIS, CCCS, CKQS
451-460: CIKS, CTDS, CKCT, CKQQ, CCIQ, CASQ, CDDY, CAPY, CADY, CPNY

(a)



(b)

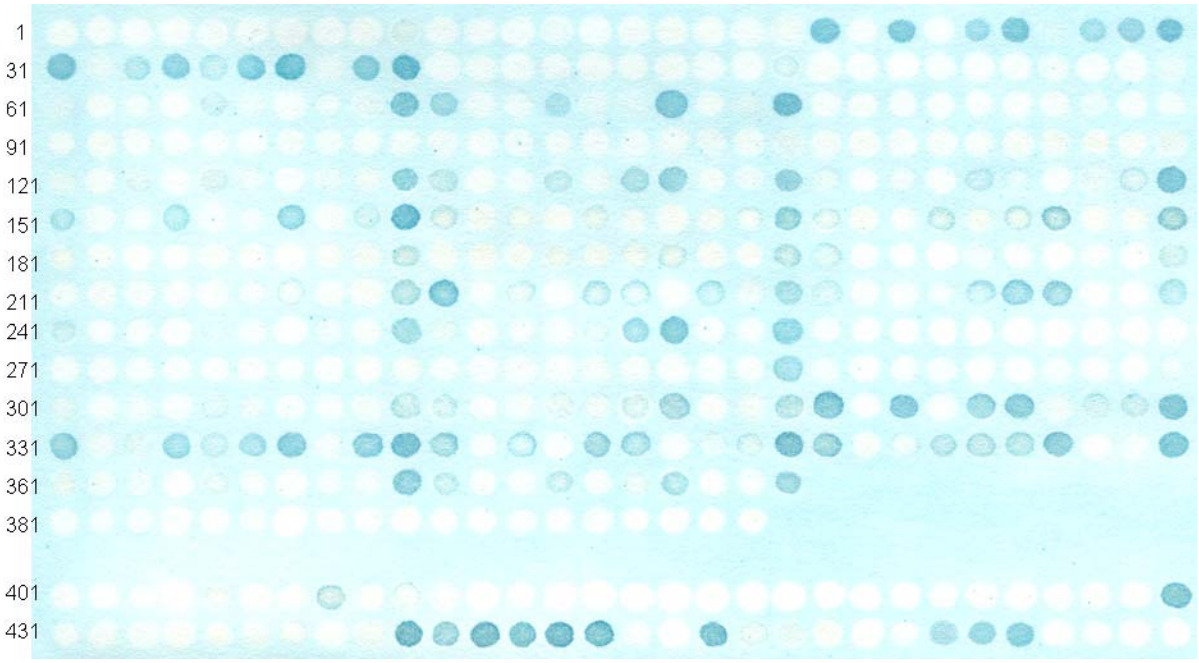


Figure S4.5 Screening of enzymatic prenylation of an RAGCVa_2X library. Each membrane was prenylated with OPP-C15-alkyne by yPFTase (2.7 ug/mL), clicked with biotin-azide and then visualized by SP-AP in BCIP solution.

Spot identity:

(a)

1-20: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)E

21-40: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)Q

41-60: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)D

61-80: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)N

81-100: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)R

101-120: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)K

121-140: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)H

141-160: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)A

161-180: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)V

181-200: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)I

201-220: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)L

221-240: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)F

241-260: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)W

261-280: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)G

281-300: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)C

301-320: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)M

321-340: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)S

341-360: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)T

361-370, 421-430: CTRK, CQRK, CKCI, CSEI, CIII, CKYI, CASL, CTIL, CSGL, CAIL

371-380, 431-440: CIIL, CIDL, CPFW, CHDE, CRME, CKGE, CCLD, CHHD, CDDD, CALD

381-390, 441-450: CWKD, CDPN, CRNR, CQGK, CLAK, CRVK, CIGK, CIHH, CYNA, CTVA

391-400, 451-460: CSNA, CPKA, CYDA, CNDV, CMYV, CFIF, CIQF, CNAG, CCVC, CSIM

401-410, 461-470: CCIM, CTIM, CIIM, CAIM, CTLM, CCPS, CDFS, CIIS, CCCS, CKQS

411-420, 471-480: CIKS, CTDS, CKCT, CKQQ, CCIQ, CASQ, CDDY, CAPY, CADY, CPNY

481-499: SVI(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,V)

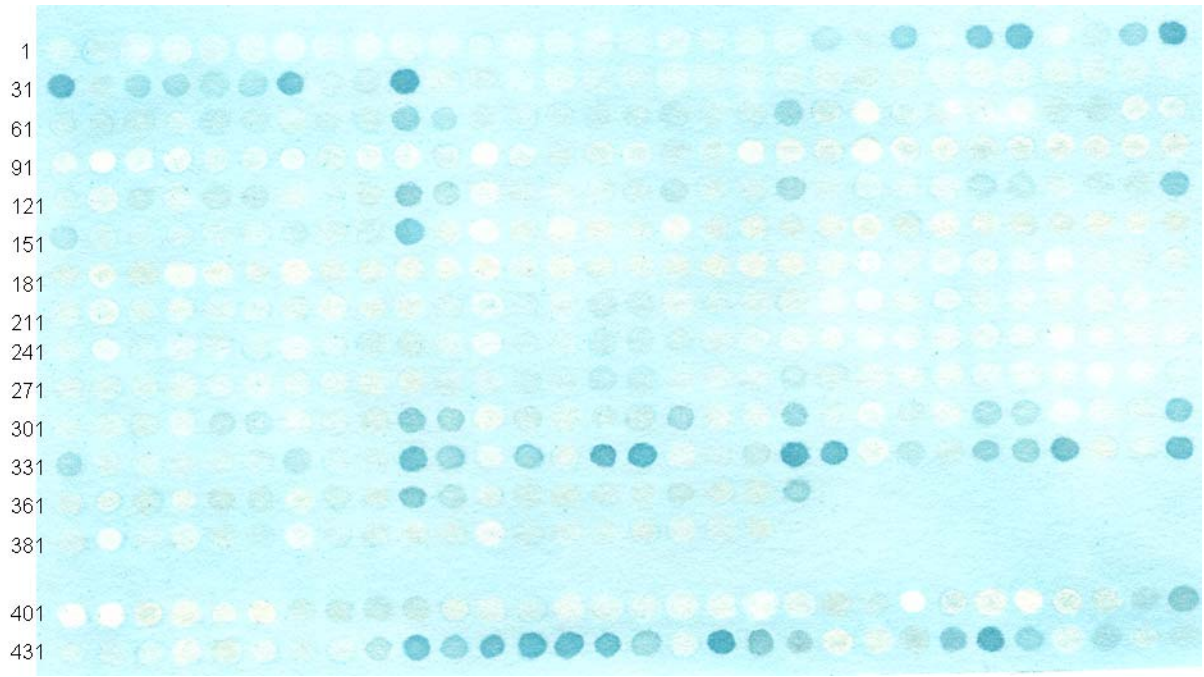
(b)

1-20: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)E

21-40: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)Q

41-60: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)D

61-80: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)N
81-100: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)R
101-120: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)K
121-140: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)H
141-160: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)A
161-180: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)V
181-200: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)I
201-220: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)L
221-240: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)F
241-260: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)Y
261-280: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)W
281-300: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)G
301-320: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)C
321-340: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)M
341-360: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)S
361-380: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)T
381-399: SVI(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,S,T,W,Y,V)
401-410: CTRK, CQRK, CKCI, CSEI, CIII, CKYI, CASL, CTIL, CSGL, CAIL
411-420: CIIL, CIDL, CPFW, CHDE, CRME, CKGE, CCLD, CHHD, CDDD, CALD
421-430: CWKD, CDPN, CRNR, CQGK, CLAK, CRVK, CIGK, CIHH, CYNA, CTVA
431-440: CSNA, CPKA, CYDA, CNDV, CMYV, CFIF, CIQF, CNAG, CCVC, CSIM
441-450: CCIM, CTIM, CIIM, CAIM, CTLM, CCPS, CDFS, CIIS, CCCS, CKQS
451-460: CIKS, CTDS, CKCT, CKQQ, CCIQ, CASQ, CDDY, CAPY, CADY, CPNY



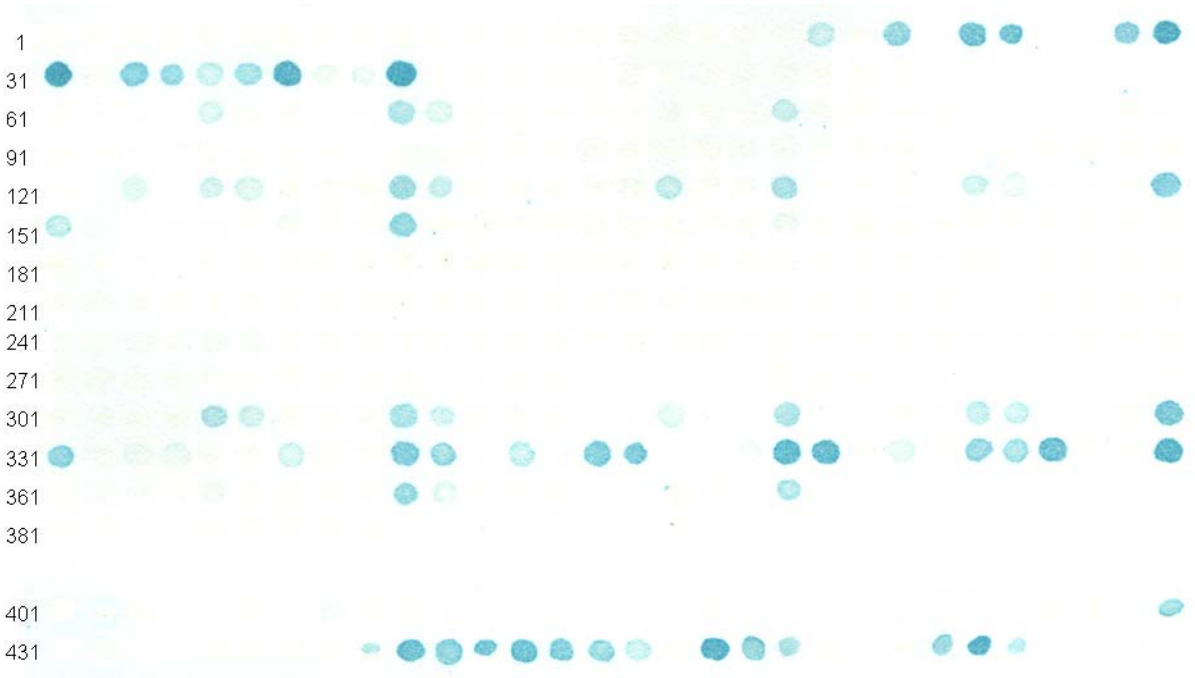


Figure S4.6 Screening of enzymatic prenylation of an RAGCVa_2X library. Each membrane was prenylated with OPP-C10-alkyne by rPFTase (40 ug/mL), clicked with biotin-azide and then visualized by SP-AP in BCIP solution.

Spot identity:

1-20: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)E

21-40: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)Q

41-60: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)D

61-80: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)N

81-100: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)R

101-120: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)K

121-140: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)H

141-160: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)A

161-180: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)V

181-200: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)I

201-220: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)L

221-240: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)F

241-260: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)Y

261-280: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)W

281-300: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)G

301-320: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)C

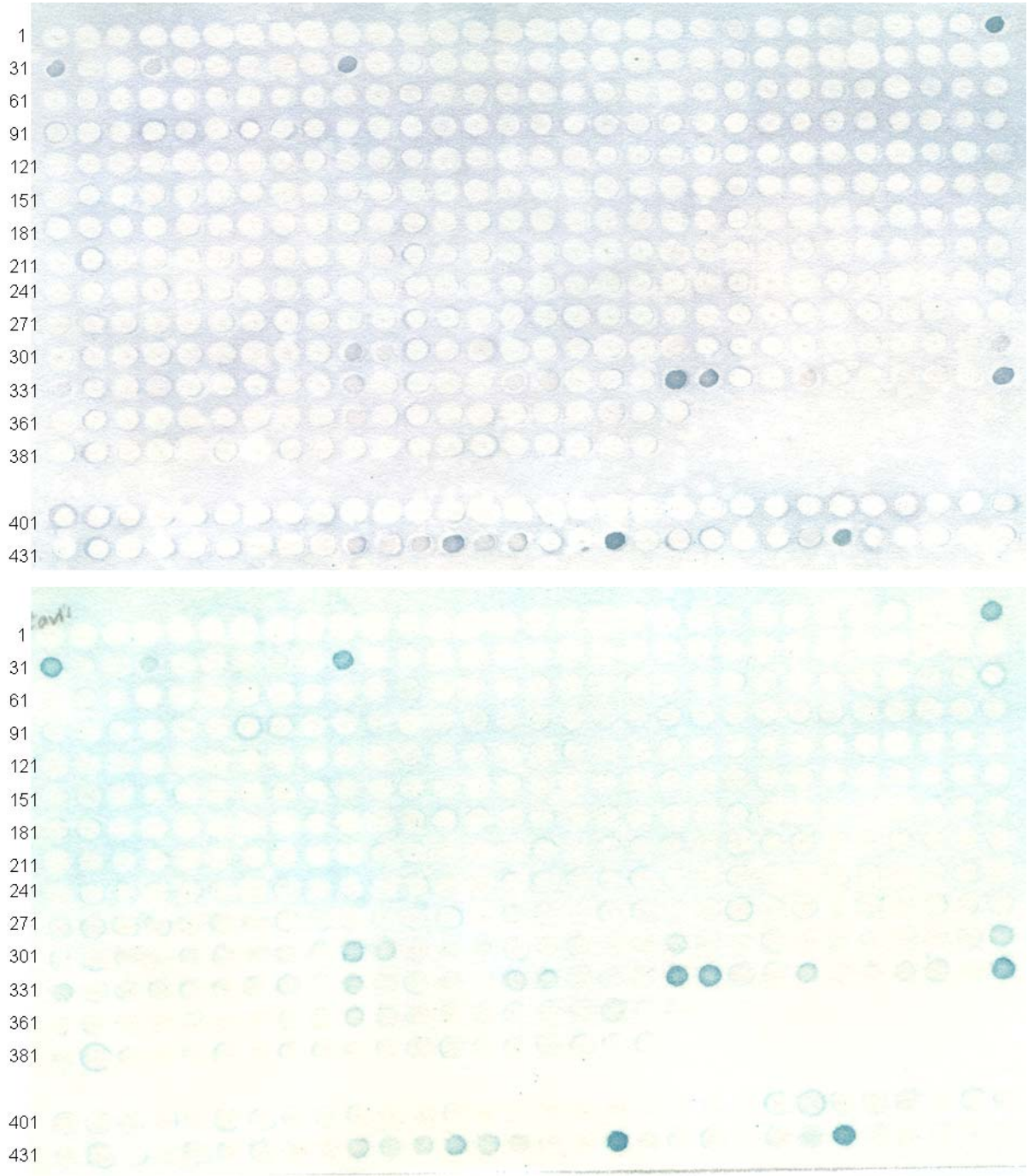
321-340: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)M

341-360: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)S

361-380: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)T

381-399: SVI(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,S,T,W,Y,V)
401-410: CTRK, CQRK, CKCI, CSEI, CIII, CKYI, CASL, CTIL, CSGL, CAIL
411-420: CIIL, CIDL, CPFW, CHDE, CRME, CKGE, CCLD, CHHD, CDDD, CALD
421-430: CWKD, CDPN, CRNR, CQGK, CLAK, CRVK, CIGK, CIHH, CYNA, CTVA
431-440: CSNA, CPKA, CYDA, CNDV, CMYV, CFIF, CIQF, CNAG, CCVC, CSIM
441-450: CCIM, CTIM, CIIM, CAIM, CTLM, CCPS, CDFS, CIIS, CCCS, CKQS
451-460: CIKS, CTDS, CKCT, CKQQ, CCIQ, CASQ, CDDY, CAPY, CADY, CPNY

(a)



(b)

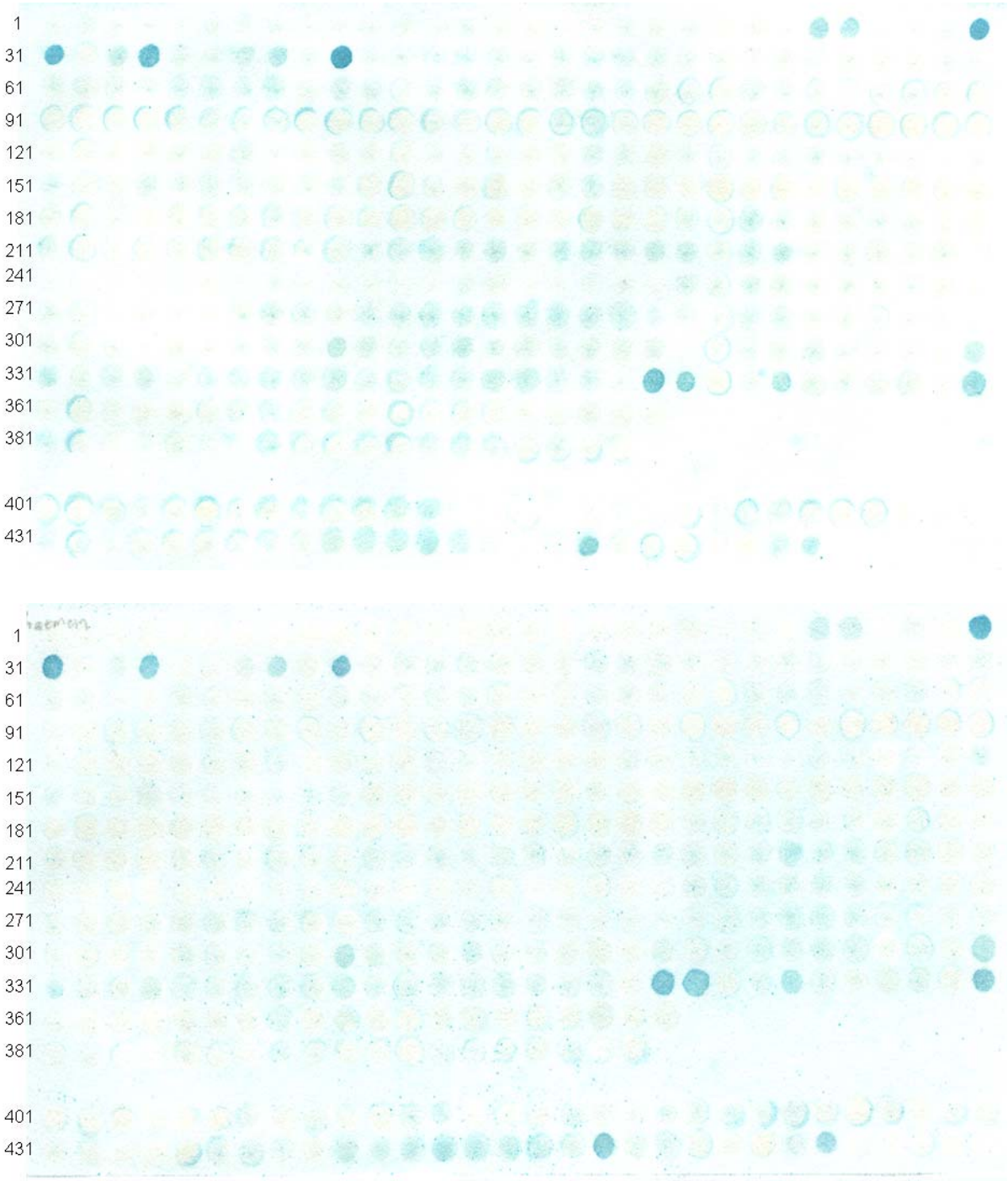


Figure S4.7 Screening of enzymatic prenylation of an RAGCVa_2X library. Each membrane was prenylated with OPP-C5-alkyne by rPFTase (55 ug/mL), clicked with biotin-azide and then visualized by SP-AP in BCIP solution. (a) The membrane is from Intavis (b) The membrane is from Whatman540

Spot identity:

1-20: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)E

21-40: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)Q

41-60: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)D
61-80: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)N
81-100: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)R
101-120: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)K
121-140: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)H
141-160: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)A
161-180: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)V
181-200: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)I
201-220: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)L
221-240: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)F
241-260: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)Y
261-280: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)W
281-300: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)G
301-320: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)C
321-340: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)M
341-360: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)S
361-380: CV(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,P,S,T,W,Y,V)T
381-399: SVI(A,R,N,D,C,Q,E,G,H,I,L,K,M,F,S,T,W,Y,V)
401-410: CTRK, CQRK, CKCI, CSEI, CIII, CKYI, CASL, CTIL, CSGL, CAIL
411-420: CIIL, CIDL, CPFW, CHDE, CRME, CKGE, CCLD, CHHD, CDDD, CALD
421-430: CWKD, CDPN, CRNR, CQGK, CLAK, CRVK, CIGK, CIHH, CYNA, CTVA
431-440: CSNA, CPKA, CYDA, CNDV, CMYV, CFIF, CIQF, CNAG, CCVC, CSIM
441-450: CCIM, CTIM, CIIM, CAIM, CTLM, CCPS, CDFS, CIIS, CCCS, CKQS
451-460: CIKS, CTDS, CKCT, CKQQ, CCIQ, CASQ, CDDY, CAPY, CADY, CPNY

(e) Kinetic analysis of individual peptides for PFTase:

(1) Comparison of rPFTase, CaPFTase and yPFTase using Ds-GCa_1a_2X peptides:

The rates of farnesylation of individual purified peptide by PFTase were determined using the time-dependent increase in fluorescence (λ_{ex} 340 nm, λ_{em} 505 nm) upon prenylation of a dansylated form of the peptide. Assays were performed with different concentrations of dansylated peptide (1, 2, 3, 5, 8 uM), 20-100 nM

PFTase, 4 μM C15-Alk, 50 mM Tris, pH 7.5, 5 mM DTT, 5 mM $MgCl_2$, 50 μM $ZnCl_2$, and 0.040 % (w/v) n-dodecyl-ß-D-maltoside at 25 °C. Reactions were initiated by the addition of PFTase. Fluorescence was measured as a function of time to define both the initial linear velocity and the reaction end point. The total fluorescence change observed upon reaction completion was divided by the initial concentration of the peptide substrate in a given reaction to yield a conversion from fluorescence units to product concentration (Amp_{conv}). The linear initial rate, V, in fluorescence intensity per minute, was then converted to a velocity (μM product produced per minute) by dividing V with Amp_{conv}. The initial rate was plotted with peptide concentration to calculate K_M and k_{cat}.

(2) Prenylation of Ds-GCVLS with C15-Alk, C10-Alk or FPP:

The rates of prenylation of purified peptides by rPFTase were determined using the time-dependent increase in fluorescence (λ_{ex} 340 nm, λ_{em} 505 nm) upon prenylation of a dansylated form of the peptide. Assays were performed with 2.0 μM dansyl-GCVLS, 20 nM rPFTase, varying concentrations (0.25-16 μM) of FPP analogues, 50 mM Tris, pH 7.5, 5 mM DTT, 5 mM $MgCl_2$, 50 μM $ZnCl_2$, and 0.040 % (*w/v*) *n*-dodecyl-*β*-D-maltoside at 25 °C. Reactions were initiated by the addition of PFTase. Fluorescence was measured as a function of time and analyzed as described above.

(3) Prenylation of OrG-RTRCVIS with C5-Alk or FPP:

Since enzymatic prenylation of dansylated peptides with C5-Alk results in minimal fluorescence increases, PFTase prenylation rates using C5-Alk were determined with an HPLC-based assay. An assay solution containing 750 uL of Tris (50 mM), pH 7.5, $MgCl_2$ (12 mm), $ZnCl_2$ (50 uM), DTT (5 mM), and OrG-RTRCVIS (2.4 mM) was prepared and allowed to incubate at RT for 20 min in order to ensure complete reduction of any disulfides. 50 uL solutions of C5-Alk were prepared so that when they were diluted into the assay solution the final concentration of C5-Alk was 5.0, 10, 20, 30, 50 or 200 uM. After the alkyne-containing solutions were added into assay buffer, rPFTase (50 uL, 8.5 uM) was added to each centrifuge tube and the reaction was incubated for 4 h at RT and then each reaction was flash frozen. Each reaction was analyzed using an HPLC equipped with a fluorescence detector. Fluorescent peaks corresponding to the starting material and product were integrated and used to calculate the concentration of prenylated product in solution. Three different samples were prepared for each concentration and the average concentration of product from the three reactions was used to calculate $K_{M,app}$ and $k_{cat,app}$ for C5-Alk.

A similar procedure was performed by using 0.2, 0.3, 0.5, 1.0, 3.0, and 20 mM concentrations of FPP. For FPP, rPFTase (50 rL, 0.85 uM) was used to initiate the reaction and the reaction was incubated for 20 min at RT before being flash frozen in

liquid nitrogen.

Reference

(1) Houghten, R. A. *Proc. Natl. Acad. Sci. USA.* **1985**, *82*, 5131.

(2) Hilpert, K.; Winkler, D. F. H.; Hancock, R. E. W. *Nat. Protoc.* **2007**, *2*, 1333.

(3) Lam, K. S.; Salmon, S. E.; Hersh, E. M.; Hruby, V. J.; Kazmierski, W. M.; Knapp, R. J. *Nature* **1991**, *354*, 82.

(4) Edman, P. *Acta Chem. Scand.* **1950**, *4*, 283.

(5) Thakkar, A.; Wavreille, A.-S.; Pei, D. *Anal. Chem.* **2006**, *78*, 5935.

(6) Biederman, K. J.; Lee, H.; Haney, C. A.; Kaczmarek, M.; Buettner, J. A. *J. Pept. Res.* **1999**, *53*, 234.

(7) Songyang, Z.; Fanning, A. S.; Fu, C.; Xu, J.; Marfatia, S. M.; Chishti, A. H.; Crompton, A.; Chan, A. C.; Anderson, J. M.; Cantley, L. C. *Science* **1997**, *275*, 73.

(8) Vaccaro, P.; Dente, L. *FEBS Lett.* **2002**, *512*, 345.

(9) Vaccaro, P.; Brannetti, B.; Montecchi-Palazzi, L.; Philipp, S.; Citterich, M. H.; Cesareni, G.; Dente, L. *J. Biol. Chem.* **2001**, *276*, 42122.

(10) Stiffler, M. A.; Chen, J. R.; Grantcharova, V. P.; Lei, Y.; Fuchs, D.; Allen, J. E.; Zaslavskaia, L. A.; MacBeath, G. *Science* **2007**, *317*, 364.

(11) Shepherd, T. R.; Hard, R. L.; Murray, A. M.; Pei, D.; Fuentes, E. J. *Biochemistry* **2011**, *50*, 1296.

(12) Takashima, S.; Kitakaze, M.; Asakura, M.; Asanuma, H.; Sanada, S.; Tashiro, F.; Niwa, H.; Miyazaki, J.-i.; Hirota, S.; Kitamura, Y.; Kitsukawa, T.; Fujisawa, H.; Klagsbrun, M.; Hori, M. *Proc. Natl. Acad. Sci. USA.* **2002**, *99*, 3657.

(13) Staton, C. A.; Kumar, I.; Reed, M. W. R.; Brown, N. J. *J. Pathol.* **2007**, *212*, 237.

(14) Becker, P. M.; Waltenberger, J.; Yachechko, R.; Mirzapoiazova, T.; Sham, J. S. K.; Lee, C. G.; Elias, J. A.; Verin, A. D. *Circ. Res.* **2005**, *96*, 1257.

(15) Acevedo, L. M.; Barillas, S.; Weis, S. M.; Göthert, J. R.; Cheresh, D. A. *Blood* **2008**, 111, 2674.

(16) von Wronski, M. A.; Raju, N.; Pillai, R.; Bogdan, N. J.; Marinelli, E. R.; Nanjappan, P.; Ramalingam, K.; Arunachalam, T.; Eaton, S.; Linder, K. E.; Yan, F.; Pochon, S.; Tweedle, M. F.; Nunn, A. D. *J. Biol. Chem.* **2006**, *281*, 5702.

(17) Starzec, A.; Ladam, P.; Vassy, R.; Badache, S.; Bouchemal, N.; Navaza, A.; du Penhoat, C. H.; Perret, G. Y. *Peptides* **2007**, *28*, 2397.

(18) Berndt, N.; Hamilton, A. D.; Sebti, S. M. *Nat. Rev. Cancer* **2011**, *11*, 775.

(19) Moores, S. L.; Schaber, M. D.; Mosser, S. D.; Rands, E.; O'Hara, M. B.; Garsky, V. M.; Marshall, M. S.; Pompliano, D. L.; Gibbs, J. B. *J. Biol. Chem.* **1991**, *266*, 14603.

(20) Fu, H. W., and Casey, P. J. *Recent Prog. Horm. Res.* **1999**, *54*, 315.

(21) Hershko, A.; Ciechanover, A. *Annu. Rev. Biochem.* **1998**, *67*, 425.

(22) Pickart, C. M. *Annu. Rev. Biochem.* **2001**, *70*, 503.

(23) Deshaies, R. J.; Joazeiro, C. A. P. *Annu. Rev. Biochem.* **2009**, *78*, 399.

(24) Wenzel, D. M.; Stoll, K. E.; Klevit, R. E. *Biochem. J.* **2010**, *433*, 31.

(25) van der Veen, A. G.; Ploegh, H. L. *Annu. Rev. Biochem.* **2012**, *81*, 323.

(26) Hicke, L.; Schubert, H. L.; Hill, C. P. *Nat. Rev. Mol. Cell. Biol.* **2005**, *6*, 610.

(27) Pai, M.-T.; Tzeng, S.-R.; Kovacs, J. J.; Keaton, M. A.; Li, S. S. C.; Yao, T.-P.; Zhou, P. *J. Mol. Biol.* **2007**, *370*, 290.

(28) Reyes-Turcu, F. E.; Horton, J. R.; Mullally, J. E.; Heroux, A.; Cheng, X.; Wilkinson, K. D. *Cell* **2006**, *124*, 1197.

(29) Williams, D. H.; Bardsley, B. *Angew. Chem. Int. Ed.* **1999**, *38*, 1172.

(30) Udugamasooriya, D. G.; Sharma, S. C.; Spaller, M. R. *ChemBioChem* **2008**, *9*, 1587.

(31) Kania, R. S.; Zuckermann, R. N.; Marlowe, C. K. *J. Am. Chem. Soc.* **1994**, *116*, 8835.

(32) Boisguerin, P.; Ay, B.; Radziwill, G.; Fritz, R. D.; Moelling, K.; Volkmer, R. *ChemBioChem* **2007**, *8*, 2302.

(33) Ay, B.; Volkmer, R.; Boisguerin, P. *Tetrahedron Lett.* **2007**, *48*, 361.

(34) Postma, T. M.; Giraud, M.; Albericio, F. *Org. Lett.* **2012**, *14*, 5468.

(35) Dekan, Z.; Mobli, M.; Pennington, M. W.; Fung, E.; Nemeth, E.; Alewood, P. F. *Angew. Chem. Int. Ed.* **2014**, *53*, 2931.

(36) Boisguerin, P.; Leben, R.; Ay, B.; Radziwill, G.; Moelling, K.; Dong, L.; Volkmer-Engert, R. *Chem. Biol.* **2004**, *11*, 449.

(37) Wiedemann, U.; Boisguerin, P.; Leben, R.; Leitner, D.; Krause, G.; Moelling, K.; Volkmer-Engert, R.; Oschkinat, H. *J. Mol. Biol.* **2004**, *343*, 703.

(38) Weski, J.; Meltzer, M.; Spaan, L.; Mönig, T.; Oeljeklaus, J.; Hauske, P.; Vouilleme, L.; Volkmer, R.; Boisguerin, P.; Boyd, D.; Huber, R.; Kaiser, M.; Ehrmann, M. *ChemBioChem* **2012**, *13*, 402.

(39) Wang, Y.-C.; Distefano, M. D. *Chem. Commun.* **2012**, *48*, 8228.

(40) Wang, Y.-C.; Dozier, J. K.; Beese, L. S.; Distefano, M. D. *ACS Chem. Biol.* **2014**.

(41) Davies, M.; Bonnat, M.; Guillier, F.; Kilburn, J. D.; Bradley, M. *The J. Org. Chem.* **1998**, *63*, 8696.

(42) Liu, R.; Marik, J.; Lam, K. S. *J. Am. Chem. Soc.* **2002**, *124*, 7678.

(43) Song, A.; Zhang, J.; Lebrilla, C. B.; Lam, K. S. *J. Am. Chem. Soc.* **2003**, *125*, 6180.

(44) Joo, S. H.; Pei, D. *Biochemistry* **2008**, *47*, 3061.

(45) Sweeney, M. C.; Pei, D. *J. Comb. Chem.* **2003**, *5*, 218.

(46) Hard, R. L.; Liu, J.; Shen, J.; Zhou, P.; Pei, D. *Biochemistry* **2010**, *49*, 10737.

(47) Yao, N.; Wu, C.-Y.; Xiao, W.; Lam, K. S. *Biopolymers* **2008**, *90*, 421.

(48) Rajagopala, S.; Casjens, S.; Uetz, P. *BMC Microbiol.* **2011**, *11*, 213. CC BY 4.0. http://creativecommons.org/licenses/by/4.0/

(49) Yang, F.; Forrer, P.; Dauter, Z.; Conway, J. F.; Cheng, N.; Cerritelli, M. E.; Steven, A. C.; Pluckthun, A.; Wlodawer, A. *Nat. Struct. Mol. Biol.* **2000**, *7*, 230.

(50) van den Berk, L. C. J.; Landi, E.; Walma, T.; Vuister, G. W.; Dente, L.; Hendriks, W. J. A. J. *Biochemistry* **2007**, *46*, 13629.

(51) Serwer, P.; Wright, E.; Hakala, K.; Weintraub, S. *BMC Res. Notes* **2008**, *1*, 36. CC BY 4.0. http://creativecommons.org/licenses/by/4.0/

(52) Kurakin, A.; Swistowski, A.; Wu, S. C.; Bredesen, D. E. *PLoS One* **2007**, *2*, e953.

(53) Sharma, S. C.; Memic, A.; Rupasinghe, C. N.; Duc, A.-C. E.; Spaller, M. R. *Biopolymers* **2009**, *92*, 183.

(54) Teesalu, T.; Sugahara, K. N.; Kotamraju, V. R.; Ruoslahti, E. *Proc. Natl. Acad. Sci. USA.* **2009**, *106*, 16157.

(55) Stricker, N. L.; Christopherson, K. S.; Yi, B. A.; Schatz, P. J.; Raab, R. W.; Dawes, G.; Bassett, D. E.; Bredt, D. S.; Li, M. *Nat. Biotech.* **1997**, *15*, 336.

(56) Palzkill, T.; Huang, W.; Weinstock, G. M. *Gene* **1998**, *221*, 79.

(57) Fuh, G.; Pisabarro, M. T.; Li, Y.; Quan, C.; Lasky, L. A.; Sidhu, S. S. *J. Biol. Chem.* **2000**, *275*, 21486.

(58) Laura, R. P.; Witt, A. S.; Held, H. A.; Gerstner, R.; Deshayes, K.; Koehler, M. F. T.; Kosik, K. S.; Sidhu, S. S.; Lasky, L. A. *J. Biol. Chem.* **2002**, *277*, 12906.

(59) Held, H. A.; Sidhu, S. S. *J. Mol. Biol.* **2004**, *340*, 587.

(60) Tonikian, R.; Zhang, Y.; Sazinsky, S. L.; Currell, B.; Yeh, J.-H.; Reva, B.; Held, H. A.; Appleton, B. A.; Evangelista, M.; Wu, Y.; Xin, X.; Chan, A. C.; Seshagiri, S.; Lasky, L. A.; Sander, C.; Boone, C.; Bader, G. D.; Sidhu, S. S. *PLoS Biol.* **2008**, *6*, e239.

(61) Ivarsson, Y.; Arnold, R.; McLaughlin, M.; Nim, S.; Joshi, R.; Ray, D.; Liu, B.; Teyra, J.; Pawson, T.; Moffat, J.; Li, S. S.-C.; Sidhu, S. S.; Kim, P. M. *Proc. Natl. Acad. Sci. USA.* **2014**, *111*, 2542.

(62) Zhao, B.; Bhuripanyo, K.; Zhang, K.; Kiyokawa, H.; Schindelin, H.; Yin, J. *Chem. Biol.* **2012**, *19*, 1265.

(63) Lee, I.; Schindelin, H. *Cell* **2008**, *134*, 268.

(64) Zhao, B.; Bhuripanyo, K.; Schneider, J.; Zhang, K.; Schindelin, H.; Boone,

D.; Yin, J. *ACS Chem. Biol.* **2012**, *7*, 2027.

(65) Zhao, B.; Zhang, K.; Villhauer, E. B.; Bhuripanyo, K.; Kiyokawa, H.; Schindelin, H.; Yin, J. *ChemBioChem* **2013**, *14*, 1323.

(66) Bos, J. L. *Cancer Res.* **1989**, *49*, 4682.

(67) Saxena, N.; Lahiri, S. S.; Hambarde, S.; Tripathi, R. P. *Cancer Invest.* **2008**, *26*, 948.

(68) Lackner, M. R.; Kindt, R. M.; Carroll, P. M.; Brown, K.; Cancilla, M. R.; Chen, C.; de Silva, H.; Franke, Y.; Guan, B.; Heuer, T.; Hung, T.; Keegan, K.; Lee, J. M.; Manne, V.; O'Brien, C.; Parry, D.; Perez-Villar, J. J.; Reddy, R. K.; Xiao, H.; Zhan, H.; Cockett, M.; Plowman, G.; Fitzgerald, K.; Costa, M.; Ross-Macdonald, P. *Cancer Cell* **2005**, *7*, 325.

(69) Rak, A.; Pylypenko, O.; Niculae, A.; Pyatkov, K.; Goody, R. S.; Alexandrov, K. *Cell* **2004**, *117*, 749.

(70) Winter-Vann, A. M.; Casey, P. J. *Nat. Rev. Cancer* **2005**, *5*, 405.

(71) Epstein, W. W.; Lever, D.; Leining, L. M.; Bruenger, E.; Rilling, H. C. *Proc. Natl. Acad. Sci. USA.* **1991**, *88*, 9668.

(72) Roskoski Jr, R. *Biochem. Biophys. Res. Commun.* **2003**, *303*, 1.

(73) Horiuchi, H.; Kawata, M.; Katayama, M.; Yoshida, Y.; Musha, T.; Ando, S.; Takai, Y. *J. Biol. Chem.* **1991**, *266*, 16981.

(74) Song, J. L.; White, T. C. *Microbiology* **2003**, *149*, 249.

(75) Chakrabarti*, D.; Azam, T.; DelVecchio, C.; Qiu, L.; Park, Y.-i.; Allen, C. M. *Mol. Biochem. Parasit.* **1998**, *94*, 175.

(76) Gelb, M. H.; Brunsveld, L.; Hrycyna, C. A.; Michaelis, S.; Tamanoi, F.; Van Voorhis, W. C.; Waldmann, H. *Nat. Chem. Biol.* **2006**, *2*, 518.

(77) Hast, M. A.; Fletcher, S.; Cummings, C. G.; Pusateri, E. E.; Blaskovich, M. A.; Rivas, K.; Gelb, M. H.; Van Voorhis, W. C.; Sebti, S. M.; Hamilton, A. D.; Beese, L. S. *Chem. Biol.* **2009**, *16*, 181.

(78) Hast, M. A.; Nichols, C. B.; Armstrong, S. M.; Kelly, S. M.; Hellinga, H. W.; Alspaugh, J. A.; Beese, L. S. *J. Biol. Chem.* **2011**, *286*, 35149.

(79) Sunami, S.; Ohkubo, M.; Sagara, T.; Ono, J.; Asahi, S.; Koito, S.; Morishima, H. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 629.

(80) Duckworth, B. P.; Xu, J.; Taton, T. A.; Guo, A.; Distefano, M. D. *Bioconjugate Chem.* **2006**, *17*, 967.

(81) Gauchet, C.; Labadie, G. R.; Poulter, C. D. *J. Am. Chem. Soc.* **2006**, *128*, 9274.

(82) Ochocki, J. D.; Distefano, M. D. *MedChemComm* **2013**, *4*, 476.

(83) Rashidian, M.; Song, J. M.; Pricer, R. E.; Distefano, M. D. *J. Am. Chem. Soc.* **2012**, *134*, 8455.

(84) Rashidian, M.; Kumarapperuma, S. C.; Gabrielse, K.; Fegan, A.; Wagner, C. R.; Distefano, M. D. *J. Am. Chem. Soc.* **2013**, *135*, 16388.

(85) Duckworth, B. P.; Chen, Y.; Wollack, J. W.; Sham, Y.; Mueller, J. D.; Taton, T. A.; Distefano, M. D. *Angew. Chem. Int. Ed.* **2007**, *46*, 8819.

(86) Liu, X.-h.; Prestwich, G. D. *J. Am. Chem. Soc.* **2002**, *124*, 20.

(87) Liu, X.-h.; Prestwich, G. D. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 2137.

(88) Owen, D. J.; Alexandrov, K.; Rostkova, E.; Scheidig, A. J.; Goody, R. S.; Waldmann, H. *Angew. Chem. Int. Ed.* **1999**, *38*, 509.

(89) Thomä, N. H.; Iakovenko, A.; Owen, D.; Scheidig, A. S.; Waldmann, H.; Goody, R. S.; Alexandrov, K. *Biochemistry* **2000**, *39*, 12043.

(90) Kim; Kleckley, T. S.; Wiemer, A. J.; Holstein, S. A.; Hohl, R. J.; Wiemer, D. F. *J. Org. Chem.* **2004**, *69*, 8186.

(91) Chen, A. P. C.; Chen, Y.-H.; Liu, H.-P.; Li, Y.-C.; Chen, C.-T.; Liang, P.-H. *J. Am. Chem. Soc.* **2002**, *124*, 15217.

(92) Dursina, B.; Reents, R.; Delon, C.; Wu, Y.; Kulharia, M.; Thutewohl, M.; Veligodsky, A.; Kalinin, A.; Evstifeev, V.; Ciobanu, D.; Szedlacsek, S. E.; Waldmann, H.; Goody, R. S.; Alexandrov, K. *J. Am. Chem. Soc.* **2006**, *128*, 2822.

(93) Das, D.; Tnimov, Z.; Nguyen, U. T. T.; Thimmaiah, G.; Lo, H.; Abankwa, D.; Wu, Y.; Goody, R. S.; Waldmann, H.; Alexandrov, K. *ChemBioChem* **2012**, *13*, 674.

(94) Chehade, K. A. H.; Kiegiel, K.; Isaacs, R. J.; Pickett, J. S.; Bowers, K. E.; Fierke, C. A.; Andres, D. A.; Spielmann, H. P. *J. Am. Chem. Soc.* **2002**, *124*, 8206.

(95) Rawat, D. S.; Krzysiak, A. J.; Gibbs, R. A. *J. Org. Chem.* **2008**, *73*, 1881.

(96) Subramanian, T.; Pais, J. E.; Liu, S.; Troutman, J. M.; Suzuki, Y.; Leela Subramanian, K.; Fierke, C. A.; Andres, D. A.; Spielmann, H. P. *Biochemistry* **2012**, *51*, 8307.

(97) Nguyen, U. T. T.; Guo, Z.; Delon, C.; Wu, Y.; Deraeve, C.; Franzel, B.; Bon, R. S.; Blankenfeldt, W.; Goody, R. S.; Waldmann, H.; Wolters, D.; Alexandrov, K. *Nat. Chem. Biol.* **2009**, *5*, 227.

(98) Scott Reid, T.; Terry, K. L.; Casey, P. J.; Beese, L. S. *J. Mol. Biol.* **2004**, *343*, 417.

(99) Kho, Y.; Kim, S. C.; Jiang, C.; Barma, D.; Kwon, S. W.; Cheng, J.; Jaunbergs, J.; Weinbaum, C.; Tamanoi, F.; Falck, J.; Zhao, Y. *Proc. Natl. Acad. Sci. USA.* **2004**, *101*, 12479.

(100) Rose, M. W.; Rose, N. D.; Boggs, J.; Lenevich, S.; Xu, J.; Barany, G.; Distefano, M. D. *J. Pept. Res.* **2005**, *65*, 529.

(101) DeGraw, A. J.; Palsuledesai, C.; Ochocki, J. D.; Dozier, J. K.; Lenevich, S.; Rashidian, M.; Distefano, M. D. *Chem. Biol. Drug Des.* **2010**, *76*, 460.

(102) Charron, G.; Tsou, L. K.; Maguire, W.; Yount, J. S.; Hang, H. C. *Mol.*

BioSyst. **2011**, *7*, 67.

(103) Charron, G.; Li, M. M. H.; MacDonald, M. R.; Hang, H. C. *Proc. Natl. Acad. Sci. USA.* **2013**, *110*, 11085.

(104) Liu, X.-h.; Suh, D.-Y.; Call, J.; Prestwich, G. D. *Bioconjugate Chem.* **2004**, *15*, 270.

(105) Lin, H. P.; Hsu, S. C.; Wu, J. C.; Sheen, I. J.; Yan, B. S.; Syu, W. J. *J. Gen. Virol.* **1999**, *80*, 91.

(106) Baron, R.; Fourcade, E.; Lajoie-Mazenc, I.; Allal, C.; Couderc, B.; Barbaras, R.; Favre, G.; Faye, J.-C.; Pradines, A. *Proc. Natl. Acad. Sci. USA.* **2000**, *97*, 11626.

(107) Troutman, J. M.; Roberts, M. J.; Andres, D. A.; Spielmann, H. P. *Bioconjugate Chem.* **2005**, *16*, 1209.

(108) Onono, F. O.; Morgan, M. A.; Spielmann, H. P.; Andres, D. A.; Subramanian, T.; Ganser, A.; Reuter, C. W. M. *Mol. Cell. Proteomics* **2010**, *9*, 742.

(109) Caplin, B. E.; Hettich, L. A.; Marshall, M. S. *Biochim. Biophys. Acta* **1994**, *1205*, 39.

(110) Pompliano, D. L.; Gomez, R. P.; Anthony, N. J. *J. Am. Chem. Soc.* **1992**, *114*, 7945.

(111) Cassidy, P. B.; Dolence, J. M.; Poulter, C. D. *Methods Enzymol.* **1995**, 250, 30.

(112) Hougland, J. L.; Hicks, K. A.; Hartman, H. L.; Kelly, R. A.; Watt, T. J.; Fierke, C. A. *J. Mol. Biol.* **2010**, *395*, 176.

(113) Parker, L. L.; Brueggemeier, S. B.; Rhee, W. J.; Wu, D.; Kent, S. B. H.; Kron, S. J.; Palecek, S. P. *Analyst* **2006**, *131*, 1097.

(114) Kress, J.; Zanaletti, R.; Amour, A.; Ladlow, M.; Frey, J. G.; Bradley, M. *Chem. Eur. J.* **2002**, *8*, 3769.

(115) Hosokawa, A.; Wollack, J.; Zhang, Z.; Chen, L.; Barany, G.; Distefano, M. *Int. J. Pept. Res. Ther.* **2007**, *13*, 345.

(116) Maurer-Stroh, S.; Koranda, M.; Benetka, W.; Schneider, G.; Sirota, F. L.; Eisenhaber, F. *PLoS Comput. Biol.* **2007**, *3*, e66.

(117) Gaon, I.; Turek, T. C.; Weller, V. A.; Edelstein, R. L.; Singh, S. K.; Distefano, M. D. *J. Org. Chem.* **1996**, *61*, 7738.

(118) Winkler, D. H.; Andresen, H.; Hilpert, K. *Methods Mol. Biol.* **2011**, 723, 105.

(119) Hartman, H. L.; Hicks, K. A.; Fierke, C. A. *Biochemistry* **2005**, *44*, 15314.

(120) Hougland, J. L.; Lamphear, C. L.; Scott, S. A.; Gibbs, R. A.; Fierke, C. A. *Biochemistry* **2009**, *48*, 1691.

(121) Reinicke, A. T.; Hutchinson, J. L.; Magee, A. I.; Mastroeni, P.; Trowsdale, J.; Kelly, A. P. *J. Biol. Chem.* **2005**, *280*, 14620.

(122) Price, C. T. D.; Al-Quadan, T.; Santic, M.; Jones, S. C.; Abu Kwaik, Y. *J. Exp. Med.* **2010**, *207*, 1713.

(123) Ivanov, S. S.; Charron, G.; Hang, H. C.; Roy, C. R. *J. Biol. Chem.* **2010**, *285*, 34686.

(124) Zeng, Q.; Si, X.; Horstmann, H.; Xu, Y.; Hong, W.; Pallen, C. J. *J. Biol. Chem.* **2000**, *275*, 21444.

(125) Nishimura, A.; Linder, M. E. *Mol. Cell. Biol.* **2013**, *33*, 1417.

(126) Yap, M. C.; Kostiuk, M. A.; Martin, D. D. O.; Perinpanayagam, M. A.; Hak, P. G.; Siddam, A.; Majjigapu, J. R.; Rajaiah, G.; Keller, B. O.; Prescher, J. A.; Wu, P.; Bertozzi, C. R.; Falck, J. R.; Berthiaume, L. G.. *J. Lipid Res.* **2010**, *51*, 1566.

(127) Hougland, J. L.; Gangopadhyay, S. A.; Fierke, C. A. *J. Biol. Chem.* **2012**, *287*, 38090.

(128) Gangopadhyay, S. A.; Losito, E. L.; Hougland, J. L. *Biochemistry* **2014**, *53*, 434.

(129) Gilleron, P.; Millet, R.; Houssin, R.; Wlodarczyk, N.; Farce, A.; Lemoine, A.; Goossens, J. F.; Chavatte, P.; Pommery, N.; Hénichart, J. P. *Eur. J. Med. Chem.* **2006**, *41*, 745.

(130) Santagada, V.; Caliendo, G.; Severino, B.; Lavecchia, A.; Perissutti, E.; Fiorino, F.; Zampella, A.; Sepe, V.; Califano, D.; Santelli, G.; Novellino, E. *J. Med. Chem.* **2006**, *49*, 1882.

(131) Maurer-Stroh, S.; Eisenhaber, F. *Genome Biol.* **2005**, *6*, R55.

(132) Krzysiak, A. J.; Rawat, D. S.; Scott, S. A.; Pais, J. E.; Handley, M.; Harrison, M. L.; Fierke, C. A.; Gibbs, R. A. *ACS Chem. Biol.* **2007**, *2*, 385.

(133) Chan, L. N.; Hart, C.; Guo, L.; Nyberg, T.; Davies, B. S. J.; Fong, L. G.; Young, S. G.; Agnew, B. J.; Tamanoi, F. *Electrophoresis* **2009**, *30*, 3598.

(134) Palsuledesai, C. C.; Ochocki, J. D.; Markowski, T. W.; Distefano, M. D. *Mol. BioSyst.* **2014**, *10*, 1094.

(135) Hannoush, R. N.; Sun, J. *Nat. Chem. Biol.* **2010**, *6*, 498.

(136) Wollack, J. W.; Silverman, J. M.; Petzold, C. J.; Mougous, J. D.; Distefano, M. D. *ChemBioChem* **2009**, *10*, 2934.